

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 21:53:36 ; Search time 508.584 Seconds
(without alignments)
1960.127 Million cell updates/sec

Title: US-09-889-611A-19

Perfect score: 23

Sequence: 1 gaatgaactacatacaaccacc 23

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_ov.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_mu.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rtd.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	23	100.0	4229	9	AF234618	AF234618 Homo sapi
2	23	100.0	157284	9	AC072051	AC072051 Homo sapi
3	23	100.0	189092	2	AC027524	AC027524 Homo sapi
4	20.4	88.7	6182	6	AX346917	AX346917 Sequence
5	18.4	80.0	687	11	BV074238	BV074238 S212P6017
6	18.4	80.0	57941	9	BX323827	BX323827 Human DNA
7	18.4	80.0	110000	2	AL928982	Continuation (5 of
8	18.4	80.0	114390	8	AP005113	AP005113 Oryza sat
9	18.4	80.0	119420	9	HS997K18	AL021406 Human DNA
10	18.4	80.0	207166	2	AC073703	AC073703 Mus muscu
11	18.4	80.0	316678	2	AC132190	AC132190 Mus muscu
12	18.2	79.1	124786	9	AC115285	AC115285 Homo sapi
13	18.2	79.1	157653	9	AC005146	AC005146 Homo sapi
14	18.2	79.1	202398	2	AC025303	AC025303 Homo sapi
15	18.2	79.1	238554	2	AC020965	AC020965 Mus muscu
16	18	78.3	37490	9	AC026356	AC026356 Homo sapi
17	18	78.3	156394	2	AC026558	AC026558 Homo sapi
18	18	78.3	215960	2	AC125959	AC125959 Rattus no
19	18	78.3	220480	2	AC023973	AC023973 Homo sapi
20	18	78.3	282132	2	AC109976	AC109976 Rattus no
21	17.8	77.4	31199	2	AC145592	AC145592 Mus muscu
22	17.8	77.4	105308	8	AF273333	AF273333 Lycopersi
23	17.8	77.4	125290	2	AC145820	AC145820 Cicer ari
24	17.8	77.4	125990	9	AC108734	AC108734 Homo sapi
25	17.8	77.4	131209	9	AC126177	AC126177 Homo sapi
26	17.8	77.4	149806	10	AC105159	AC105159 Mus muscu
27	17.8	77.4	152898	2	AC134432	AC134432 Mus muscu
28	17.8	77.4	162446	9	AL137118	AL137118 Human DNA
29	17.8	77.4	163012	2	AC142104	AC142104 Mus muscu
30	17.8	77.4	164603	2	AC119466	AC119466 Rattus no
31	17.8	77.4	167170	10	AL669968	AL669968 Mouse DNA
32	17.8	77.4	167229	2	AC101950	AC101950 Mus muscu
33	17.8	77.4	176278	10	AC126557	AC126557 Mus muscu
34	17.8	77.4	181060	5	BX119907	BX119907 Zebrafish
35	17.8	77.4	182116	9	AC093831	AC093831 Homo sapi
36	17.8	77.4	185211	8	OSJN00158	AL662955 Oryza sat
37	17.8	77.4	187532	9	AC008333	AC008333 Homo sapi
38	17.8	77.4	192092	10	AC126248	AC126248 Mus muscu
39	17.8	77.4	193856	2	BX537353	BX537353 Danio rer
40	17.8	77.4	194778	2	AC118844	AC118844 Rattus no
41	17.8	77.4	196265	10	AC116325	AC116325 Mus muscu
42	17.8	77.4	198338	10	AC087098	AC087098 Genomic s
43	17.8	77.4	211236	2	AC120182	AC120182 Mus muscu
44	17.8	77.4	220320	2	AC132527	AC132527 Rattus no
45	17.8	77.4	228428	10	AL596456	AL596456 Mouse DNA

ALIGNMENTS

RESULT 1
AF234618 4229 bp DNA linear PRI 29-MAR-2002
LOCUS Homo sapiens meggin gene, promoter region and partial sequence.
DEFINITION AF234618
ACCESSION AF234618.2 GI:19808130
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 4229)
AUTHORS Miyata, T., Nangaku, M., Inagi, R. and Kurokawa, K.
TITLE Transcriptional regulation of a mesangium-predominant gene, meggin
JOURNAL Unpublished


```

/clone_lib="RPC1-11 Human Male BAC"
599. .635
/rpt_family="("T)"
1040. .1077
/rpt_family="AT_rich"
1411. .1747
/rpt_family="LIP"
2632. .2992
/rpt_family="MLT1A1"
3115. .3138
/rpt_family="AT_rich"
/rpt_family="LIMC4a"
3834. .4338
/rpt_family="MLT1C"
complement(4339. .5397)
/rpt_family="LIMC4a"
complement(5742. .5899)
/rpt_family="MER5B"
6499. .6816
/rpt_family="HAL1"
complement(7416. .7621)
/rpt_family="MER3"
complement(8163. .9090)
/rpt_family="L2"
9276. .9419
/rpt_family="MIR"
9598. .9658
/rpt_family="MER81"
9659. .10033
/rpt_family="MLT1J"
13985. .14009
/rpt_family="AT_rich"
14011. .14115
/rpt_family="(TA)"
complement(14942. .15044)
/rpt_family="MIR3"
15155. .15298
/rpt_family="L2"
15592. .15773
/rpt_family="LIM4c"
15874. .18310
/rpt_family="LIP1A1c"
16373. .16374
/Note="1328 bp of bacterial transposon insertion in BAC
excised between these 2 bp"
18311. .18331
/rpt_family="(TAAAAA)"
18332. .19566
/rpt_family="LIP1A1c"
19567. .19587
/rpt_family="(CA)"
19588. .20042
/rpt_family="LIP1A1c"
20043. .20064
/rpt_family="AT_rich"
complement(21259. .21386)
/rpt_family="MIR3"
22395. .22420
/rpt_family="AT_rich"
complement(22788. .22948)
/rpt_family="FRAM"
complement(23843. .24253)
/rpt_family="MSTA"
complement(24678. .25041)
/rpt_family="MLT1A2"
25054. .25094
/rpt_family="HAL1"
25214. .25358
/rpt_family="LIM4c"
complement(25602. .25854)
/rpt_family="MIR"
26745. .26796

```

```

repeat_region /rpt_family="AT_rich"
27451. .27484
/rpt_family="(CAAAA)"
complement(28183. .28219)
/rpt_family="MIR"
complement(28249. .28325)
/rpt_family="MIR"
complement(29372. .29459)
/rpt_family="MIR"
29539. .30003
/rpt_family="MER66B"
30596. .30624
/rpt_family="AT_rich"
31418. .31829
/rpt_family="L2"
31846. .31915
/rpt_family="(TGAA)"
complement(32019. .32103)

Query Match 100.0%; Score 23; DB 9; Length 157284;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATGAACCTACATACACACCACC 23
DB 155067 GAATGAACCTACATACACACCACC 155045

RESULT 3
AC027524 189092 bp DNA linear HTG 27-MAR-2003
LOCUS Homo sapiens chromosome 18 clone RP11-317G1 map 18, 3 unordered
DEFINITION pieces.
AC027524
VERSION AC027524.4 GI:21307437
KEYWORDS HTG; HTGS PHASE1; HTGS_FULLTOP; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189092)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-317G1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189092)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lechoczky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Piarre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
TITLE
JOURNAL
REFERENCE 30-MAR-2000 Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS 3 (bases 1 to 189092)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,

```

Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, J., Lehotzky, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menees, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phukhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, K., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Tallamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.

Direct Submission
Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 1, 2002 this sequence version replaced gi:11990731.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8516
Center clone name: 317_G_1

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 85659: contig of 85659 bp in length
* 85660 85759: gap of 100 bp
* 85760 163014: contig of 77255 bp in length
* 163015 163114: gap of 100 bp
* 163115 189092: contig of 25978 bp in length.

----- Location/Qualifiers
1. 189092
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-317G1"
/clone_lib="RPC1-11 Human Male BAC"

ORIGIN

Query Match 100.0%; Score 23; DB 2; Length 189092;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAATGAACACTACATACACACC 23

Db 66563 GAATGAACACTACATACACACC 66585

RESULT 4

AX346917/c
LOCUS
DEFINITION Sequence 1988 from Patent WO0200928.
ACCESSION AX346917

AX346917.1 GI:18494803

KEYWORDS

SOURCE
ORGANISM
synthetic construct
artificial sequences.

REFERENCE

1 Olek, A., Pispembrock, C. and Berlin, K.
Diagnosis of diseases associated with the immune system
Patent: WO 0200928-A 1988 03-JAN-2002;
JOURNAL
Epigenomics AG (DE)

FEATURES

source
1. 6182
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 88.7%; Score 20.4; DB 6; Length 6182;
Best Local Similarity 95.5%; Pred. No. 55;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AATGAACACTACATACACACC 23

Db 1310 AATAAATACTACATACACACC 1289

RESULT 5

AX346917.1
LOCUS
DEFINITION S21260172F88.T0 CZECHII/Ei Mus musculus STS genomic, sequence
tagged site.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)

1 687
GI:31190033

REFERENCE

1 (bases 1 to 687)
Wade, C.M., Kulbokas, E.J. III, Kirby, A.W., Zody, M.C., Mullikin, J.C., Lander, E.S., Lindblad-Toh, K., and Daly, M.J.
The mosaic structure of variation in the laboratory mouse genome
Nature 420 (6915), 574-578 (2002)
22354684
PUBMED
12466852

AUTHORS

Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersli@genome.wi.mit.edu
Primer A: None
Primer B: None
STS size: 687
Protocol:

COMMENT

WGS-discovery: Paired-end low-coverage whole genome shotgun reads
were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS
reads were placed uniquely on the MGSV3 C57BL/6J assembly and SNP
detection was carried out by SSAHA-SNP. 225,000 reads were
annotated
as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
and the strain from which the particular read came. The validation
rate for these SNPs was estimated at approximately 98%.

FEATURES

source
1. 687
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="CZECHII/Ei"
/db_xref="taxon:10090"
/map="- 12 22-639 20493381-20492764"


```

STS
ORIGIN
Query Match 80.0%; Score 18.4; DB 11; Length 687;
Best Local Similarity 95.0%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AATGAACACTACATACACCA 21
|||||
Db 495 AATGAACACTACATACACCA 514

RESULT 6
BX323827 57941 bp DNA linear PRI 29-JUL-2003
LOCUS Human DNA sequence from clone RPA-732K23 on chromosome X, complete
DEFINITION
ACCESSION BX323827
VERSION BX323827.5 GI:33341976
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Whitehead,S.
Direct Submission
Submitted (29-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jul 29, 2003 this sequence version replaced gi:32169142.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the rare
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
En:, EMBL; SW:, SWISSPROT; Tr:, TrEMBL; Wp:, WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep/ This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RPA-732K23 is from the library RPCI-4 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.
FEATURES
source
1. 57941
/mol_type="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP4-732K23"
/clone_lib="RPCI-4"

/clone_lib="CZCHII/E1"
<1..5687

Query Match 80.0%; Score 18.4; DB 11; Length 687;
Best Local Similarity 95.0%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AATGAACACTACATACACCA 21
|||||
Db 51313 AATGATCTACATACACCA 51332

RESULT 7
AL928982 4
WPCOMMENT
Sequence split into 7 fragments LOCUS AL928982 Accession AL928982
Fragment Name Begin End
AL928982_0 1 110000
AL928982_1 10001 210000
AL928982_2 20001 310000
AL928982_3 30001 410000
AL928982_4 40001 510000
AL928982_5 50001 610000
AL928982_6 60001 671561
Continuation (5 of 7) of AL928982 from base 400001 (AL928982 Homo sapiens chromosome X c
Query Match 80.0%; Score 18.4; DB 2; Length 110000;
Best Local Similarity 95.0%; Pred. No. 3.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AATGACTACATACACCA 21
|||||
Db 97907 AATGATCTACATACACCA 97926

RESULT 8
AP005113 114390 bp DNA linear PLN 12-SEP-2003
LOCUS Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 2,
DEFINITION PAC clone:P0685G12, complete sequence.
ACCESSION AP005113
VERSION AP005113.2 GI:34740250
KEYWORDS HTG.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1
AUTHORS Sasaki, T., Matsumoto, T. and Katayose, Y.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, PAC
clone:P0685G12
JOURNAL Published Only in Database (2002)
REFERENCE
2 (bases 1 to 114390)
AUTHORS Sasaki, T., Matsumoto, T. and Katayose, Y.
TITLE Direct Submission
JOURNAL Submitted (18-APR-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT On Sep 15, 2003 this sequence version replaced gi:20219001.
The orientation of the sequence is from SP6 to 17 of the PAC clone.
FEATURES
source
1. 114390
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="2"
/clone="P0685G12"

ORIGIN
Query Match 80.0%; Score 18.4; DB 8; Length 114390;

```



```

repeat_region /note="AluX repeat: matches 39. .312 of consensus"
31226. .31518
repeat_region /note="AluJo repeat: matches 1. .299 of consensus"
32108. .32413
repeat_region /note="AluX repeat: matches 1. .306 of consensus"
32743. .32894
repeat_region /note="MIR repeat: matches 50. .213 of consensus"
33424. .33614
misc_feature /note="match: GSS: Em:B91428"
34293. .34333
misc_feature /note="match: STS: Em:G09780"
complement(34293. 34333)
misc_feature /note="match: STS: Em:G09898"
34303. .34332
repeat_region /note="10 copies 3 mer tta 100% conserved"
34334. .34676
repeat_region /note="LIP47 repeat: matches 5807. .6145 of consensus"
34716. .34947
repeat_region /note="LINE repeat: matches 5576. .5810 of consensus"
35064. .35135
repeat_region /note="18 copies 4 mer tata 60% conserved"
35065. .35132
repeat_region /note="34 copies 2 mer at 67% conserved"
36181. .36570
repeat_region /note="LIR16B repeat: matches 38. .458 of consensus"
36610. .36986
repeat_region /note="L2 repeat: matches 1596. .2008 of consensus"
37799. .37860
repeat_region /note="MLT11 repeat: matches 74. .135 of consensus"
38576. .38734
repeat_region /note="LJMA9 repeat: matches 5855. .5989 of consensus"
38802. .38913
repeat_region /note="LMA9 repeat: matches 6163. .6270 of consensus"
40678. .40811
repeat_region /note="LMA2 repeat: matches 6175. .6308 of consensus"
40812. .41083
repeat_region /note="LJMC4 repeat: matches 7710. .7970 of consensus"
42665. .42718
repeat_region /note="LIP3 repeat: matches 5501. .5554 of consensus"
43334. .43362
repeat_region /note="MLT11 repeat: matches 112. .140 of consensus"
44826. .45020
repeat_region /note="MLT1A1 repeat: matches 37. .233 of consensus"
45035. .45382
repeat_region /note="LMD3 repeat: matches 7330. .7676 of consensus"
45383. .45737
repeat_region /note="THE1C repeat: matches 1. .371 of consensus"
45738. .45796
repeat_region /note="LMD3 repeat: matches 7272. .7330 of consensus"
46215. .47939
repeat_region /note="LIP45 repeat: matches 4407. .6141 of consensus"
48417. .48504
repeat_region /note="22 copies 4 mer acag 72% conserved"
48461. .48504
repeat_region /note="22 copies 2 mer ac 90% conserved"
49354. .49574
repeat_region /note="MIR repeat: matches 5. .230 of consensus"
50009. .50539
repeat_region /note="LMA4 repeat: matches 3878. .4424 of consensus"
51611. .51907
repeat_region /note="AluX repeat: matches 1. .300 of consensus"
52619. .52658
repeat_region /note="20 copies 2 mer ca 90% conserved"
52970. .53161

```

Query Match 80.0%; Score 18.4; DB 9; Length 119420;
 Best Local Similarity 95.0%; Pred. No. 3.7e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AATGAACATACATACACCA 21

DB 7849 AATGAACATACATACACCA 7830

RESULT 10
 AC073703
 LOCUS
 DEFINITION

AC073703
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

AC073703 207166 bp DNA linear HTG 18-JUL-2000
 Mus musculus clone RP23-16E5, WORKING DRAFT SEQUENCE, 22 ordered
 pieces.

AC073703
 AC073703.2 GI:9256762
 HTG; HTGS_PHASE2; HTGS_DRAFT.
 Mus musculus (house mouse)

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 207166)
 DOE Joint Genome Institute.
 Sequencing of Mouse

Unpublished
 2 (bases 1 to 207166)
 DOE Joint Genome Institute.

Direct Submission
 Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Jul 18, 2000 this sequence version replaced gi:8810320.

-----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: http://www.jgi.doe.gov

 Project Information
 Center Project Name: 1743117
 Center clone name: RPCI-23_16E5

 Summary Statistics
 Consensus quality: 193525 bases at least Q40
 Consensus quality: 202548 bases at least Q30
 Consensus quality: 204188 bases at least Q20
 Estimated insert size: 208000; agarose-fp estimation
 Estimated insert size: 206166; sum-of-contigs estimation
 Quality coverage: 8.62 in Q20 bases; agarose-fp estimation
 Quality coverage: 8.7 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 22 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

1 4743: contig of 4743 bp in length
 4744 4843: gap of unknown length
 13587: contig of 8744 bp in length
 13588 13687: gap of unknown length
 26532: contig of 12845 bp in length
 26533 26632: gap of unknown length
 26633 39968: contig of 13336 bp in length
 40068: gap of unknown length
 60987: contig of 20929 bp in length
 61098: gap of unknown length
 71379: contig of 10282 bp in length
 71380 71479: gap of unknown length
 93270: contig of 21791 bp in length
 93271 93370: gap of unknown length
 93371 99451: contig of 6081 bp in length
 99452 109513: contig of 9962 bp in length
 109514 109613: gap of unknown length
 109614 116641: contig of 7028 bp in length
 116642 128540: contig of 11799 bp in length
 128541 128640: gap of unknown length
 128641 135192: contig of 6552 bp in length
 135193 135292: gap of unknown length

135293 137898: contig of 2606 bp in length
 * 137899 137998: gap of unknown length
 * 143409 143408: contig of 5410 bp in length
 * 143409 143508: gap of unknown length
 * 143509 145880: contig of 2372 bp in length
 * 145881 145980: gap of unknown length
 * 145981 161155: contig of 15175 bp in length
 * 161156 161253: gap of unknown length
 * 161254 165223: contig of 3868 bp in length
 * 165224 165223: gap of unknown length
 * 165224 167221: contig of 1998 bp in length
 * 167222 167321: gap of unknown length
 * 167322 191494: contig of 24173 bp in length
 * 191495 191595: gap of unknown length
 * 191596 192681: contig of 1087 bp in length
 * 192682 192781: gap of unknown length
 * 192782 203772: contig of 10991 bp in length
 * 203773 203872: gap of unknown length
 * 203873 207166: contig of 3294 bp in length.
 * 207167 Location/Qualifiers
 1. 207166
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone_lib="RP23-16E5"
 /clone_lib="RPCI mouse BAC library 23"
 ORIGIN
 Query Match 80.0%; Score 18.4; DB 2; Length 207166;
 Best Local Similarity 95.0%; Pred. No. 3.5e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 2 AATGAACCTACATACACCA 21
 |||||
 Db 93079 AATGACCTACATACACCA 93098
 |||||
 RESULT 11
 AC132190/c
 LOCUS
 DEFINITION Mus musculus clone RP23-378B3, WORKING DRAFT SEQUENCE, 41 unordered
 pieces.
 AC132190
 VERSION AC132190.2 GI:30017734
 KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
 SOURCE Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Birren,B., Nussbaum,C. and Lander,E.
 1. (bases 1 to 316678)
 TITLE
 JOURNAL
 REFERENCE 2 (bases 1 to 316678)
 AUTHORS
 Birren,B., Nussbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
 Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fero,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., Mlenka,V.,
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission

Submitted (30-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 316678)
 Birren,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
 Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collymore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Fero,S.,
 Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Liu,A., Mabbitt,R., MacLean,C.,
 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
 Meldrim,J., Meneus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J.,
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 17, 2003 this sequence version replaced gi:22549775.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L26973
 Center clone name: 378_B_3

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 41 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 906: contig of 906 bp in length
 * 907 1006: gap of 100 bp
 * 1007 1622: contig of 616 bp in length
 * 1623 1722: gap of 100 bp
 * 1723 2454: contig of 732 bp in length
 * 2455 2554: gap of 100 bp
 * 2555 3462: contig of 908 bp in length
 * 3463 3562: gap of 100 bp
 * 3563 4941: contig of 1378 bp in length
 * 4942 5040: gap of 100 bp
 * 5041 6356: contig of 1316 bp in length
 * 6357 6456: gap of 100 bp
 * 6457 7539: contig of 1083 bp in length
 * 7540 7639: gap of 100 bp
 * 7640 8951: contig of 1312 bp in length
 * 8952 9052: gap of 100 bp
 * 9053 10202: contig of 1151 bp in length
 * 10203 10302: gap of 100 bp
 * 10303 11848: contig of 1546 bp in length
 * 11849 11948: gap of 100 bp
 * 11949 13055: contig of 1107 bp in length
 * 13056 13156: gap of 100 bp
 * 13157 15527: contig of 2372 bp in length
 * 15528 15627: gap of 100 bp
 * 15628 17409: contig of 1782 bp in length
 * 17410 17509: gap of 100 bp

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

```

* 17510 18310: contig of 801 bp in length
* 18311 18410: gap of 100 bp
* 18411 20473: contig of 2063 bp in length
* 20474 20573: gap of 100 bp
* 20574 22133: contig of 1560 bp in length
* 22134 22233: gap of 100 bp
* 22234 24255: contig of 2022 bp in length
* 24256 24356: gap of 100 bp
* 24357 25980: contig of 1625 bp in length
* 25981 26081: gap of 100 bp
* 26082 27721: contig of 1641 bp in length
* 27722 27821: gap of 100 bp
* 27822 30199: contig of 2378 bp in length
* 30200 32155: gap of 100 bp
* 32156 32255: gap of 100 bp
* 32256 34250: contig of 1995 bp in length
* 34251 34350: gap of 100 bp
* 34351 36191: contig of 1841 bp in length
* 36192 36291: gap of 100 bp
* 36292 39231: contig of 2940 bp in length
* 39232 39331: gap of 100 bp
* 39332 43671: contig of 4340 bp in length
* 43672 43771: gap of 100 bp
* 43772 45876: contig of 2105 bp in length
* 45877 45976: gap of 100 bp
* 45977 48712: contig of 2736 bp in length
* 48713 48812: gap of 100 bp
* 48813 52480: contig of 3668 bp in length
* 52481 52580: gap of 100 bp
* 52581 56188: contig of 3608 bp in length
* 56189 56288: gap of 100 bp
* 56289 59456: contig of 3168 bp in length
* 59457 59556: gap of 100 bp
* 59557 63722: contig of 4166 bp in length
* 63723 63822: gap of 100 bp
* 63823 67556: contig of 3734 bp in length
* 67557 67656: gap of 100 bp
* 67657 72152: contig of 4496 bp in length
* 72153 72252: gap of 100 bp
* 72253 74581: contig of 2329 bp in length
* 74582 74681: gap of 100 bp
* 74682 79961: contig of 5280 bp in length
* 79962 80061: gap of 100 bp
* 80062 85912: contig of 5851 bp in length
* 85913 86013: gap of 100 bp
* 86013 96395: contig of 10383 bp in length
* 96396 96495: gap of 100 bp
* 96496 142316: contig of 45821 bp in length
* 142317 142416: gap of 100 bp
* 142417 201149: contig of 58733 bp in length
* 201150 201249: gap of 100 bp
* 201250 263398: contig of 62149 bp in length
* 263399 263498: gap of 100 bp
* 263499 316678: contig of 53180 bp in length.
FEATURES
    Location/Qualifiers
        1..316678
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /clone="RP23-378B3"
        /clone_lib="RPCI-23 Female Mouse BAC"
misc_feature 1..906
    /note="assembly_fragment"
    clone_end:SP6
    vector_side:left
misc_feature 1007..1622
    /note="assembly_fragment"
misc_feature 1723..2454
    /note="assembly_fragment"
misc_feature 2555..3462
    /note="assembly_fragment"
misc_feature 3563..4940

```

```

/note="assembly_fragment"
5041..6356
/note="assembly_fragment"
6457..7539
/note="assembly_fragment"
7640..8951
/note="assembly_fragment"
9052..10202
/note="assembly_fragment"
10303..11848
/note="assembly_fragment"
11949..13055
/note="assembly_fragment"
13156..15527
/note="assembly_fragment"
15628..17409
/note="assembly_fragment"
17510..18310
/note="assembly_fragment"
18411..20473

Query Match      80.0%; Score 18.4; DB 2; Length 316678;
Best Local Similarity 95.0%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2  AATGAACACTACATCAACCA 21
Db      223361 AATGAACACTACAAACCA 223342
|||||
AC115285/c
AC115285
LOCUS      AC115285      124786 bp      DNA      linear      PRI 29-JUN-2002
DEFINITION Homo sapiens chromosome 1 clone RP4-580019, complete sequence.
ACCESSION   AC115285 AL359872
VERSION     AC115285.2 GI:21629400
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 124786)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
TITLE        Direct Submission
JOURNAL      Unpublished
REFERENCE   2 (bases 1 to 124786)
AUTHORS     Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE        Direct Submission
JOURNAL      Submitted (16-MAR-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3 (bases 1 to 124786)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
TITLE        Direct Submission
JOURNAL      Submitted (29-JUN-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Jun 29, 2002 this sequence version replaced gi:19526124.
            ----- Genome Center
            Center: University of Washington
            Center Code: UWGC
            Web site: http://www.genome.washington.edu
            Contact: uwgchtgs@u.washington.edu
            Drafting Center: SC
            ----- Project Information
            Center project name: chr-1
            Center clone name: RP4-580019 (sc0808)
            ----- Summary Statistics
            Sequencing vector: plasmid; 41% of reads
            Sequencing vector: plasmid; 108752; 59% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads

```

Assembly program: Phrap; version 0.990319
 Consensus quality: 124541 bases at least Q40
 Consensus quality: 124757 bases at least Q30
 Consensus quality: 124780 bases at least Q20
 Insert size: 124786; sum-of-contigs
 Quality coverage: 6.2x in Q20 bases; sum-of-contigs

Overlapping Sequences:

5: RPL-149P10 AL138837, 17366-bp overlap
 3: RPS-947L8 AL355178, 46965-bp overlap

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

NsiI				BglII				HindIII			
SeqDerMap	FrgrPrnt	SeqDerMap	FrgrPrnt	SeqDerMap	FrgrPrnt	SeqDerMap	FrgrPrnt	SeqDerMap	FrgrPrnt	SeqDerMap	FrgrPrnt
8243	8134	13702	14292	3464	3373						
3357	3404	5671	5558	449	<800						
266	<800	4488	4326	512	<800						
1629	1613	10240	9842	2814	2849						
9716	9788	2889	3110	1247	1222						
5663	5484	2673	2659	4695	4771						
6253	6313	597	<800	3373	3373						
5429	5484	1973	1984	4630	4595						
10024	9788	3382	3602	1454	1438						
5452	5484	2476	2449	9110	8939						
3066	3090	154	<800	1751	1762						
354	<800	184	<800	5564	5531						
5599	5484	765	757	546	<800						
5566	5484	3932	3853	4849	4771						

FEATURES

source
 Location/Qualifiers
 1..124786
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP4-580019"
 /clone_lib="RP4 human PAC library 4"
 1..4
 /note="Single subclone region"
 40666..40934
 /note="Single subclone region"

misc_feature

misc_feature

```

misc_feature 42390..42459
/note="Single subclone region"
misc_feature 60247..60254
/note="Single subclone region"
misc_feature 124398..124468
/note="Single subclone region"
misc_feature 124783..124786
/note="Single subclone region"

ORIGIN
Query Match 79.1%; Score 18.2; DB 9; Length 124786;
Best Local Similarity 87.0%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAATGACTACATACACACC 23
Db 33813 GATGGAAGTACATACACACC 33791

RESULT 13
AC005146
LOCUS
DEFINITION Homo sapiens 12 PAC RPI-157K6 (Roswell Park Cancer Institute Human
ACCESSION AC005146
VERSION complete sequence.
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 157653)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbosa, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burrell, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S.,
Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P.,
Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N.,
Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P.,
Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M.,
Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S.,
Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y.,
Johnson, R., Jolivet, S., Joudah, S., Karisson, E., Kelly, S., Khan, U.,
King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N.,
Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O.,
Lien, C., Liu, J., Liu, W., Loulsegged, H., Lozadó, R.J., Lu, X.,
Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P.,
Marondel, I., Martin, R., Martindale, A., Martinez, E., Massey, E.,
Mawney, E., McLeod, M.P., Meador, M., Mei, G., Merscher, S.,
Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T.,
Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M.,
Neal, D., Nelson, D., Newkirk, J., Newton, N., Nguyen, S., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenko, S., Ogih, M., Okwum, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pichens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shim, C., Shoshitani, N., Sisson, I.,
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vaquer, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Kucherlapati, R.,

```

```

TITLE
JOURNAL
REFERENCE
2 (bases 1 to 157653)
Worley, K.C.
Direct Submission
Submitted (20-JUN-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 157653)
Worley, K.C.
Direct Submission
Submitted (29-AUG-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 157653)
Worley, K.C.
Direct Submission
Submitted (01-SEP-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 157653)
Worley, K.C.
Direct Submission
Submitted (03-SEP-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 157653)
Worley, K.C.
Direct Submission
Submitted (17-SEP-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
7 (bases 1 to 157653)
Worley, K.C.
Direct Submission
Submitted (11-NOV-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
8 (bases 1 to 157653)
Worley, K.C.
Direct Submission
Submitted (10-JAN-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
9 (bases 1 to 157653)
Worley, K.C.
Direct Submission
Submitted (20-FEB-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Aug 29, 1998 this sequence version replaced gi:3327895.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

```

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES

source

Location/Qualifiers

1. .157653

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="12"

/clone="RP1-157K6"

complement(1..2001)

/note="overlaps bases 1..2001 of clone AC004806"

/function="clone overlap"

74..284

/rpt_family="MER82"

complement(1961..2045)

/rpt_family="L2"

3408..3501

/rpt_family="MIR"

3602..3653

/rpt_family="MIR"

3698..3847

/rpt_family="MIR"

3872..4164

/rpt_family="AluY"

4225..4754

/rpt_family="MLTIE2"

4784..5062

/rpt_family="MLTIE2-internal"

5123..5286

/rpt_family="MLTIE2-internal"

5674..6777

/rpt_family="L1M4c"

complement(6835..7243)

/rpt_family="L1P2"

7344..7530

/rpt_family="L1MD3"

7688..8058

/rpt_family="MER113"

8371..8497

/rpt_family="MIR"

complement(10516..11013)

/rpt_family="MLTIE2"

14787..14807

/rpt_family="AT-rich"

16143..16381

/standard_name="D128340"

16196..16315

/standard_name="D128340"

16260..16292

/rpt_family="(TG)n"

17707..17926

/standard_name="A007A34"

17815..17945

/standard_name="SGC31838"

18920..19201

/rpt_family="AluSq"

21082..21265

/standard_name="RH93075"

22757..23045

/rpt_family="AluSx"

repeat_region 25096..25120

/rpt_family="AT-rich"

Query Match 79.1%; Score 18.2; DB 9; Length 157653;

Best Local Similarity 87.0%; Pred. No. 4.4e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAATGAACATACATACACACC 23

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

RESULT 14

AC025303

LOCUS

DEFINITION

Homo sapiens clone RP11-689C19, WORKING DRAFT SEQUENCE, 33

unordered pieces.

AC025303

AC025303.1 GI:7209939

HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 202398)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

REFERENCE

2 (bases 1 to 202398)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., LaRocque,K., Lamarez,R., Landers,T., Lehoczy,J.,

Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,

Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,

Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,

Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,

Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4340

Center clone name: 689C_19

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 174122 bases at least Q40

Consensus quality: 186313 bases at least Q30

Consensus quality: 192877 bases at least Q20

Insert size: 194000; agarose-fp

Insert size: 199198; sum-of-contigs

Quality coverage: 3.5 in Q20 bases; agarose-fp
 Quality coverage: 3.4 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 33 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1395: contig of 1395 bp in length
* 1396 1495: gap of 100 bp
* 1496 2752: contig of 1257 bp in length
* 2753 2852: gap of 100 bp
* 2853 4422: contig of 1570 bp in length
* 4423 4522: gap of 100 bp
* 4523 6760: contig of 2238 bp in length
* 6761 6860: gap of 100 bp
* 6861 8709: contig of 1849 bp in length
* 8710 8809: gap of 100 bp
* 8810 11245: contig of 2436 bp in length
* 11246 11345: gap of 100 bp
* 11346 13448: contig of 2103 bp in length
* 13449 13548: gap of 100 bp
* 13549 15972: contig of 2424 bp in length
* 15973 16072: gap of 100 bp
* 16073 19092: contig of 3020 bp in length
* 19093 19192: gap of 100 bp
* 19193 22045: contig of 2853 bp in length
* 22046 22145: gap of 100 bp
* 22146 25821: contig of 3676 bp in length
* 25822 25921: gap of 100 bp
* 25922 30209: contig of 4288 bp in length
* 30210 30309: gap of 100 bp
* 30310 33625: contig of 3316 bp in length
* 33626 33725: gap of 100 bp
* 33726 38386: contig of 4661 bp in length
* 38387 38486: gap of 100 bp
* 38487 43134: contig of 4648 bp in length
* 43135 43234: gap of 100 bp
* 43235 47752: contig of 4518 bp in length
* 47753 47852: gap of 100 bp
* 47853 52841: contig of 4989 bp in length
* 52842 52941: gap of 100 bp
* 52942 58193: contig of 5252 bp in length
* 58194 58294: gap of 100 bp
* 58294 63615: contig of 5321 bp in length
* 63616 63715: gap of 100 bp
* 63715 69041: contig of 5327 bp in length
* 69042 75114: gap of 100 bp
* 75115 75214: contig of 5973 bp in length
* 75215 82031: contig of 100 bp
* 82032 82131: gap of 100 bp
* 82132 87882: contig of 5751 bp in length
* 87883 95376: gap of 100 bp
* 95377 95476: contig of 7394 bp in length
* 95477 102941: contig of 100 bp
* 102942 103041: gap of 100 bp
* 103042 111576: contig of 8535 bp in length
* 111577 111676: gap of 100 bp
* 111677 120340: contig of 8664 bp in length
* 120341 120441: gap of 100 bp
* 120442 131444: contig of 11004 bp in length
* 131445 131544: gap of 100 bp
* 131545 143190: contig of 11646 bp in length
* 143191 143290: gap of 100 bp
* 143291 156345: contig of 13055 bp in length
* 156346 156445: gap of 100 bp
* 156446 169908: contig of 13463 bp in length
* 169909 170008: gap of 100 bp
    
```

```

* 170009 186176: contig of 16168 bp in length
* 186177 186276: gap of 100 bp
* 186277 202398: contig of 16122 bp in length.
FEATURES
    Location/Qualifiers
        1..202398
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /clone="RP11-689C19"
            /clone_lib="RPC1-11 Human Male BAC"
        1..1395
            /note="assembly_fragment"
        1496..2752
            /note="assembly_fragment"
        2853..4422
            /note="assembly_fragment"
        4523..6760
            /note="assembly_fragment"
        6861..8709
            /note="assembly_fragment"
            clone_end:T7
            vector_side:left
        8810..11245
            /note="assembly_fragment"
        11346..13448
            /note="assembly_fragment"
        13549..15972
            /note="assembly_fragment"
        16073..19092
            /note="assembly_fragment"
        19193..22045
            /note="assembly_fragment"
        22146..25821
            /note="assembly_fragment"
        25922..30209
            /note="assembly_fragment"
        30310..33625
            /note="assembly_fragment"
        33726..38386
            /note="assembly_fragment"
        38487..43134
            /note="assembly_fragment"
        43235..47752
            /note="assembly_fragment"
        47853..52841
            /note="assembly_fragment"
        52942..58193
            /note="assembly_fragment"
        58294..63614
            /note="assembly_fragment"
        63715..69041
            /note="assembly_fragment"
        69142..75114
            /note="assembly_fragment"
        75215..82031
            /note="assembly_fragment"
        82132..87882
            /note="assembly_fragment"
        87983..95376
            /note="assembly_fragment"
        95477..102941
            /note="assembly_fragment"
        103042..111576
            /note="assembly_fragment"
        111677..120340
            /note="assembly_fragment"
        120441..131444
            /note="assembly_fragment"
    
```

Query Match 79.1%; Score 18.2; DB 2; Length 202398;
 Best Local Similarity 87.0%; Pred. No. 4.3e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;


```

* * 72154 75289: gap of unknown length
* * contig of 3136 bp in length
* * 75290 78389: gap of unknown length
* * contig of 3100 bp in length
* * 78390 80910: gap of unknown length
* * contig of 2521 bp in length
* * 80911 82153: gap of unknown length
* * contig of 1243 bp in length
* * 82154 84007: gap of unknown length
* * contig of 1854 bp in length
* * 84008 86443: gap of unknown length
* * contig of 2436 bp in length
* * 86444 89640: gap of unknown length
* * contig of 3197 bp in length
* * 89641 91024: gap of unknown length
* * contig of 1384 bp in length
* * 91025 93347: gap of unknown length
* * contig of 2323 bp in length
* * 93348 95487: gap of unknown length
* * contig of 2140 bp in length
* * 95488 98489: gap of unknown length
* * contig of 3002 bp in length
* * 98490 100499: gap of unknown length
* * contig of 2010 bp in length
* * 100500 104757: gap of unknown length
* * contig of 4258 bp in length
* * 104758 106505: gap of unknown length
* * contig of 1748 bp in length
* * 106506 112258: gap of unknown length
* * contig of 5753 bp in length
* * 112259 115111: gap of unknown length
* * contig of 2853 bp in length
* * 115112 117906: gap of unknown length
* * contig of 2795 bp in length
* * 117907 120203: gap of unknown length
* * contig of 2297 bp in length
* * 120204 125155: gap of unknown length
* * contig of 4952 bp in length
* * 125156 132035: gap of unknown length
* * contig of 6880 bp in length
* * 132036 138866: gap of unknown length
* * contig of 6831 bp in length
* * 138867 146355: gap of unknown length
* * contig of 7489 bp in length
* * 146356 155442: gap of unknown length
* * contig of 9087 bp in length
* * 155443 161962: gap of unknown length
* * contig of 6520 bp in length
* * 161963 170957: gap of unknown length
* * contig of 8995 bp in length
* * 170958 181490: gap of unknown length
* * contig of 10533 bp in length
* * 181491 193728: gap of unknown length
* * contig of 12238 bp in length
* * 193729 204122: gap of unknown length
* * contig of 10394 bp in length
* * 204123 223034: gap of unknown length
* * contig of 18912 bp in length
* * 223035 238554: gap of unknown length
* * contig of 15520 bp in length.
FEATURES Location/Qualifiers
Query Match 79.1%; Score 18.2; DB 2; Length 238554;
Best Local Similarity 87.0%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GAATGAACATACATACACC 23
Db 9289 GAATGAACATACATACACC 9311

```

Search completed: March 24, 2004, 23:01:21
 Job time : 515.584 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 18:56:32 ; Search time 120.427 Seconds
(without alignments)
811.351 Million cell updates/sec

Title: US-09-889-611A-19

Perfect score: 23

Sequence: 1 gaatgaactacatacaaccacc 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	23	100.0	23	3	AAA71452 Human meg
2	23	100.0	1431	3	AAA71435 Human meg
3	22	95.7	30	3	AAA71448 Human meg
4	22	95.7	128	3	AAA71434 Human meg
5	20.4	88.7	6182	6	ABL34015 Human imm
6	17.2	74.8	350	5	AAS34210 Human cdn
7	17.2	74.8	580	6	ABQ23445 Oligonuc
8	17.2	74.8	580	6	ABQ23444 Oligonuc
9	17.2	74.8	1350	5	AAS72599 DNA encod
10	17.2	74.8	4789	9	ADD47985 Rat gene
11	17.2	74.8	9087	6	ABK31277 Signal tr
12	17.2	74.8	9087	6	ABL70238 Chemi call
13	17.2	74.8	9087	6	AAS61181 Human gen
14	17.2	74.8	12879	6	ABK92230 Prostata
15	17.2	74.8	13202	4	AAK51828 Human pol
16	17.2	74.8	16983	9	ADD47173 Rat gene
17	17	73.9	16918	6	ABL33617 Human imm
18	16.8	73.0	744	7	ACF69209 Photorhab
19	16.8	73.0	1215	7	ACA39407 Prokaryot
20	16.8	73.0	2477	4	ABL27492 Drosophil
21	16.8	73.0	2815	4	AAK91396 Human col
22	16.8	73.0	2815	4	AAI57764 Human col
23	16.8	73.0	2815	6	ABS99941 Genomic D

24	16.8	73.0	2815	9	ADB93094	ADB93094 Human col
25	16.8	73.0	2851	4	AAK91397	AAK91397 Human dig
26	16.8	73.0	2851	4	AAK91395	AAK91395 Human dig
27	16.8	73.0	2851	4	AAI57763	AAI57763 Human col
28	16.8	73.0	2851	4	AAI57765	AAI57765 Human col
29	16.8	73.0	2851	6	ABS99940	ABS99940 Genomic D
30	16.8	73.0	2851	6	ABS99942	ABS99942 Genomic D
31	16.8	73.0	2851	9	ADB93093	ADB93093 Human col
32	16.8	73.0	2851	9	ADB93095	ADB93095 Human col
33	16.8	73.0	8262	4	AAK71151	AAK71151 Human imm
34	16.8	73.0	9365	6	ABL33526	ABL33526 Human imm
35	16.8	73.0	110000	2	AAT58840_1	Continuation (2 of
36	16.8	73.0	110000	7	ACF67367_21	Continuation (22 o
37	16.8	73.0	243072	7	ACF65382	ACF65382 Photorhab
38	16.6	72.2	219	3	AAC22299	AAC22299 Human sec
39	16.6	72.2	390	7	ACA48438	ACA48438 Prokaryot
40	16.6	72.2	461	7	ABZ55960	ABZ55960 Aspergill
41	16.6	72.2	664	4	AAI20886	AAI20886 Human bre
42	16.6	72.2	4569	4	ABL15968	ABL15968 Drosophil
43	16.6	72.2	17993	6	AAK46721	AAK46721 Human tra
44	16.6	72.2	24978	2	AAK60209	AAK60209 SEQ ID 3
45	16.6	72.2	24979	3	AAA52321	AAA52321 Genomic D

ALIGNMENTS

RESULT 1
AAA71452
ID AAA71452 standard; DNA; 23 BP.
XX
AC AAA71452;
XX
DT 01-DEC-2000 (first entry)
XX
DE Human megsin promoter PCR primer SEQ ID NO: 19.
XX
KW Promoter; megsin; human; protein isolation; screening. PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200043528-A1.
XX
PD 27-JUL-2000.
XX
PF 25-JAN-2000; 2000WO-JP000350.
XX
PR 25-JAN-1999; 99JP-00015667.
XX
PA (KURO//) KUROKAWA K.
PA (MIYA//) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 2000-543257/49.
XX
PS DNA for promoter region of megsin useful for screening proteins.
XX
CC Example 4; Page 42; 45pp; Japanese.
XX
CC This invention describes a novel DNA sequence (I) representing a promoter
CC region having part or all of a specific base sequence. The invention also
CC describes (1) a vector containing (I); (2) a cell transformed by the
CC above vector; and (3) protein produced using (I). (I) is useful for
CC screening and isolating proteins (especially transcription factors).
CC AAA71434-A71469 represent PCR primers used in the method described in the
CC invention
XX
SQ Sequence 23 BP; 11 A; 7 C; 2 G; 3 T; 0 U; 0 Other;

Query Match 100.0%; Score 23; DB 3; Length 23;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GAATGAAGTACATACCAACACC 23
DB 1 GAATGAAGTACATACCAACACC 23

RESULT 2
AAA71435
ID AAA71435 standard; DNA; 1431 BP.
AC AAA71435;
XX
XX 01-DEC-2000 (first entry)
XX Human meggin promoter fragment DNA.
XX Promoter; meggin; human; protein isolation; screening. ss.
XX Homo sapiens.
XX WO200043528-A1.
XX
XX 27-JUL-2000.
XX
XX 25-JAN-2000; 2000WO-JP000350.
XX
XX 25-JAN-1999; 99JP-00015667.
XX
XX (KURO/) KUROKAWA K.
XX (MIYA/) MIYATA T.
XX Miyata T;
XX
XX WPI; 2000-543257/49.
XX
XX DNA for promoter region of meggin useful for screening proteins.
XX Example 5; Page 40; 45pp; Japanese.
XX
XX This invention describes a novel DNA sequence (I) representing a promoter
XX region having part or all of a specific base sequence. The invention also
XX describes (1) a vector containing (I); (2) a cell transformed by the
XX above vector; and (3) protein produced using (I). (I) is useful for
XX screening and isolating proteins (especially transcription factors).
XX AAA71434-A71469 represent PCR primers used in the method described in the
XX invention
XX
XX Sequence 30 BP; 13 A; 8 C; 3 G; 6 T; 0 U; 0 Other;
Query Match 95.7%; Score 22; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AATGAAGTACATACCAACACC 23
DB 1 AATGAAGTACATACCAACACC 22

RESULT 4
AAA71434
ID AAA71434 standard; DNA; 128 BP.
AC AAA71434;
XX
XX 01-DEC-2000 (first entry)
XX Human meggin promoter fragment DNA.
XX Promoter; meggin; human; protein isolation; screening. ss.
XX Homo sapiens.
XX WO200043528-A1.
XX
XX 27-JUL-2000.
XX
XX 25-JAN-2000; 2000WO-JP000350.
XX
XX 25-JAN-1999; 99JP-00015667.
XX
XX (KURO/) KUROKAWA K.
XX (MIYA/) MIYATA T.
XX Miyata T;
XX
XX WPI; 2000-543257/49.
XX
XX DNA for promoter region of meggin useful for screening proteins.
XX Claim 1; Page 32; 45pp; Japanese.
XX
XX This invention describes a novel DNA sequence (I) representing a promoter
XX region having part or all of a specific base sequence. The invention also
XX describes (1) a vector containing (I); (2) a cell transformed by the

```

CC above vector; and (3) protein produced using (1). (1) is useful for
 CC screening and isolating proteins (especially transcription factors). This
 CC sequence represents the human meglin promoter which is described in the
 CC method of the invention

SQ Sequence 128 BP; 44 A; 29 C; 24 G; 29 T; 0 U; 2 Other;
 Query Match 95.7%; Score 22; DB 3; Length 128;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AATGAACACTACATAACACACC 23
 DB 1 AATGAACACTACATAACACACC 22

RESULT 5
 ABL34015/C
 ID ABL34015 standard; DNA; 6182 BP.
 AC ABL34015;
 DT 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 1988.
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytostatic; neutropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 ds.

OS Homo sapiens.
 XX WO200200928-A2.
 PN 03-JAN-2002.
 XX 02-JUL-2001; 2001WO-EP007537.
 XX 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 PA (EPIG-) EPIGENOMICS AG.
 XX Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-130909/17.
 DR Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.

XX Claim 1; SEQ ID NO 1988; 32pp + Sequence Listing; German.
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX Sequence 6182 BP; 1934 A; 37 C; 1059 G; 3152 T; 0 U; 0 Other;

SQ Query Match 88.7%; Score 20.4; DB 6; Length 6182;
 Best Local Similarity 95.5%; Pred. No. 22;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AATGAACACTACATAACACACC 23
 DB 1310 AATAAACTACATAACACACC 1289
 RESULT 6
 AAS34210/C
 ID AAS34210 standard; cDNA; 350 BP.
 XX AAS34210;
 AC AAS34210;
 DT 17-DEC-2001 (first entry)
 DE Human cDNA encoding a novel foetal antigen, SEQ ID No 734.
 KW Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;
 KW immunomodulator; cardiovascular; cytostatic; nephrothropic;
 KW cardiovascular; autoimmune disease; rheumatoid arthritis;
 KW hyperproliferative disorder; breast neoplasm; cancer;
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischaemia; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; epithelial cell proliferation; food additive.
 XX Homo sapiens.
 OS WO200155312-A2.
 PN 02-AUG-2001.
 XX 17-JAN-2001; 2001WO-US001321.
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.

```
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232196P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232406P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246603P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249219P.
PR 17-NOV-2000; 2000US-0249220P.
PR 17-NOV-2000; 2000US-0249221P.
PR 17-NOV-2000; 2000US-0249222P.
PR 17-NOV-2000; 2000US-0249223P.
PR 17-NOV-2000; 2000US-0249224P.
PR 17-NOV-2000; 2000US-0249225P.
PR 17-NOV-2000; 2000US-0249226P.
PR 17-NOV-2000; 2000US-0249227P.
PR 17-NOV-2000; 2000US-0249228P.
PR 17-NOV-2000; 2000US-0249229P.
PR 17-NOV-2000; 2000US-0249230P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251988P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254037P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-488782/53.
XX P-PSDB; AAU21390.
XX
XX New polynucleotides and polypeptides for diagnosing, treating, preventing
or prognosing e.g. diseases or disorders of the nervous, musculoskeletal,
excretory, gastrointestinal, reproductive, and respiratory systems.
XX
XX Claim 1; SEQ ID NO 734; 642pp; English.
XX
XX The invention relates to novel nucleic acids encoding novel human foetal
antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
in diagnosing a pathological condition or susceptibility to a
pathological condition. The antibodies to the antigens can also be used
in alleviating symptoms associated with the disorders and in diagnostic
immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
(ELISA). Disorders which are diagnosed or treated include autoimmune
diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
nervous system disorders e.g. Alzheimer's disease, infections caused by
bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
The polypeptides can also be used to aid wound healing and epithelial
cell proliferation, to prevent skin aging due to sunburn, to maintain
organs before transplantation, for supporting cell culture of primary
tissues, to regenerate tissues and in chemotaxis. The polypeptides can
also be used as a food additive or preservative to increase or decrease
storage capabilities, fat content, lipid, protein, carbohydrate, Numerous
vitamins, minerals, cofactors and other nutritional components. Numerous
examples of diseases and disorders treated by the nucleic acids and
proteins are given in the specification. The present sequence encodes a
fetal antigen of the invention. Note: The sequence data for this patent
did not form part of the printed specification, but was obtained in
Query Match 74.8%; Score 17.2; DB 5; Length 350;
Best Local Similarity 86.4%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AATGAACCTACATACACCACC 23
DB 145 AATGAACCTCATTACATCACC 124
RESULT 7
ABQ23445
```

```

ID  ABQ23445 standard; DNA; 580 BP.
XX  ABQ23445;
XX  12-JUL-2002 (first entry)
XX  Oligonucleotide for detecting cytosine methylation SEQ ID NO 10036.
XX  Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX  drug; side effect; cancer; central nervous system; cardiovascular;
XX  gastrointestinal; respiratory system; single nucleotide polymorphism;
XX  SNP; cell differentiation; ds.
XX  Homo sapiens.
XX  WO200218632-A2.
XX  07-MAR-2002.
XX  01-SEP-2001; 2001WO-EP010074.
XX  01-SEP-2000; 2000DE-01043826.
XX  05-SEP-2000; 2000DE-01044543.
XX  (EPIG-) EPIGENOMICS AG.
XX  Olek A, Piepenbrock C, Berlin K, Guetig D;
XX  WPI; 2002-371829/40.
XX  Determining the degree of cytosine methylation in genomic DNA, useful for
XX  diagnosis and prognosis, comprises selective hybridization of amplicons
XX  from chemically treated DNA.
XX  Claim 12; 56pp + Sequence Listing; 56pp; German.
XX  This invention describes a novel method for determining the degree of
XX  methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX  genomic sample of DNA. The sample is treated chemically to convert
XX  cytosine (C) but not methylated C, to uracil, then part of the genomic
XX  DNA that contains the target C is amplified to form a labeled amplicon.
XX  The amplicon is hybridised to two classes, each with at least one member,
XX  of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX  degree of hybridisation to both classes is determined from the label on
XX  the amplicon. From the ratio of labels hybridised to the two classes of
XX  oligomers, the degree of methylation is calculated. The method is used:
XX  (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX  and of a wide range of diseases, e.g. cancer, disorders of the central
XX  nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX  particularly by detecting mutations or single nucleotide polymorphisms
XX  (SNP's); and (ii) for differentiation of cell or tissue types and for
XX  investigating cell differentiation. The method allows the methylation
XX  status of many C residues to be determined simultaneously. ABQ13410-
XX  ABQ54121 represent genomic DNA sequences used to illustrate the method
XX  for determining the degree of cytosine methylation described in the
XX  disclosure of the invention
XX  Sequence 580 BP; 256 A; 150 C; 49 G; 125 T; 0 U; 0 Other;
XX  Query Match 74.8%; Score 17.2; DB 6; Length 580;
XX  Best Local Similarity 86.4%; Pred. No. 4.8e+02;
XX  Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX  QY 2 AATGAACACTACATCAACGCC 23
XX  Db 254 AATCAACTACATATCAACGCC 275
XX  RESULT 8
XX  ABQ23444/c
XX  ID ABQ23444 standard; DNA; 580 BP.
XX  AC ABQ23444;
XX  12-JUL-2002 (first entry)
XX  Oligonucleotide for detecting cytosine methylation SEQ ID NO 10036.
XX  Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX  drug; side effect; cancer; central nervous system; cardiovascular;
XX  gastrointestinal; respiratory system; single nucleotide polymorphism;
XX  SNP; cell differentiation; ds.
XX  Homo sapiens.
XX  WO200218632-A2.
XX  07-MAR-2002.
XX  01-SEP-2001; 2001WO-EP010074.
XX  01-SEP-2000; 2000DE-01043826.
XX  05-SEP-2000; 2000DE-01044543.
XX  (EPIG-) EPIGENOMICS AG.
XX  Olek A, Piepenbrock C, Berlin K, Guetig D;
XX  WPI; 2002-371829/40.
XX  Determining the degree of cytosine methylation in genomic DNA, useful for
XX  diagnosis and prognosis, comprises selective hybridization of amplicons
XX  from chemically treated DNA.
XX  Claim 12; 56pp + Sequence Listing; 56pp; German.
XX  This invention describes a novel method for determining the degree of
XX  methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX  genomic sample of DNA. The sample is treated chemically to convert
XX  cytosine (C) but not methylated C, to uracil, then part of the genomic
XX  DNA that contains the target C is amplified to form a labeled amplicon.
XX  The amplicon is hybridised to two classes, each with at least one member,
XX  of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX  degree of hybridisation to both classes is determined from the label on
XX  the amplicon. From the ratio of labels hybridised to the two classes of
XX  oligomers, the degree of methylation is calculated. The method is used:
XX  (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX  and of a wide range of diseases, e.g. cancer, disorders of the central
XX  nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX  particularly by detecting mutations or single nucleotide polymorphisms
XX  (SNP's); and (ii) for differentiation of cell or tissue types and for
XX  investigating cell differentiation. The method allows the methylation
XX  status of many C residues to be determined simultaneously. ABQ13410-
XX  ABQ54121 represent genomic DNA sequences used to illustrate the method
XX  for determining the degree of cytosine methylation described in the
XX  disclosure of the invention
XX  Sequence 580 BP; 256 A; 150 C; 49 G; 125 T; 0 U; 0 Other;
XX  Query Match 74.8%; Score 17.2; DB 6; Length 580;
XX  Best Local Similarity 86.4%; Pred. No. 4.8e+02;
XX  Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX  QY 2 AATGAACACTACATCAACGCC 23
XX  Db 254 AATCAACTACATATCAACGCC 275
XX  RESULT 8
XX  ABQ23444/c
XX  ID ABQ23444 standard; DNA; 580 BP.
XX  AC ABQ23444;

```

```

XX  12-JUL-2002 (first entry)
XX  Oligonucleotide for detecting cytosine methylation SEQ ID NO 10035.
XX  Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX  drug; side effect; cancer; central nervous system; cardiovascular;
XX  gastrointestinal; respiratory system; single nucleotide polymorphism;
XX  SNP; cell differentiation; ds.
XX  Homo sapiens.
XX  WO200218632-A2.
XX  07-MAR-2002.
XX  01-SEP-2001; 2001WO-EP010074.
XX  01-SEP-2000; 2000DE-01043826.
XX  05-SEP-2000; 2000DE-01044543.
XX  (EPIG-) EPIGENOMICS AG.
XX  Olek A, Piepenbrock C, Berlin K, Guetig D;
XX  WPI; 2002-371829/40.
XX  Determining the degree of cytosine methylation in genomic DNA, useful for
XX  diagnosis and prognosis, comprises selective hybridization of amplicons
XX  from chemically treated DNA.
XX  Claim 12; 56pp + Sequence Listing; 56pp; German.
XX  This invention describes a novel method for determining the degree of
XX  methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX  genomic sample of DNA. The sample is treated chemically to convert
XX  cytosine (C) but not methylated C, to uracil, then part of the genomic
XX  DNA that contains the target C is amplified to form a labeled amplicon.
XX  The amplicon is hybridised to two classes, each with at least one member,
XX  of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX  degree of hybridisation to both classes is determined from the label on
XX  the amplicon. From the ratio of labels hybridised to the two classes of
XX  oligomers, the degree of methylation is calculated. The method is used:
XX  (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX  and of a wide range of diseases, e.g. cancer, disorders of the central
XX  nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX  particularly by detecting mutations or single nucleotide polymorphisms
XX  (SNP's); and (ii) for differentiation of cell or tissue types and for
XX  investigating cell differentiation. The method allows the methylation
XX  status of many C residues to be determined simultaneously. ABQ13410-
XX  ABQ54121 represent genomic DNA sequences used to illustrate the method
XX  for determining the degree of cytosine methylation described in the
XX  disclosure of the invention
XX  Sequence 580 BP; 125 A; 49 C; 150 G; 256 T; 0 U; 0 Other;
XX  Query Match 74.8%; Score 17.2; DB 6; Length 580;
XX  Best Local Similarity 86.4%; Pred. No. 4.8e+02;
XX  Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX  QY 2 AATGAACACTACATCAACGCC 23
XX  Db 327 AATCAACTACATATCAACGCC 306
XX  RESULT 9
XX  AAS72599/c
XX  ID AAS72599 standard; cDNA; 1350 BP.
XX  AC AAS72599;
XX  DT 13-FEB-2002 (first entry)
XX

```


DE DNA encoding novel human diagnostic protein #8403.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR P-FSDB; ABG08412.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 1; SEQ ID NO 8403; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1350 BP; 374 A; 400 C; 171 G; 404 T; 0 U; 1 Other;
Query Match 74.8%; Score 17.2; DB 5; Length 1350;
Best Local Similarity 86.4%; Pred. No. 5.1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AATGAACACTACATACACCACC 23
|||||
Db 1018 AATGATCTACACAAACCACC 997
RESULT 10
ADD47985/c
ID ADD47985 standard; DNA; 4789 BP.
XX AC ADD47985;
XX DT 29-JAN-2004 (first entry)
XX DE Rat gene U30813, SEQ ID NO 13681.
XX

KW Rat; ds; gene; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
XX spared nerve injury; SNI; Chung.
XX Rattus norvegicus.
XX WO2003016475-A2.
XX PD 27-FEB-2003.
XX PF 14-AUG-2002; 2002WO-US025765.
XX PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.
XX PR 26-NOV-2001; 2001US-0333347P.
XX (GEHO) GEN HOSPITAL CORP.
XX PA (FARB) BAYER AG.
XX PI Woolf C, D'urso D, Befort K, Costigan M;
XX DR WPI; 2003-268312/26.
XX DR GENBANK; U30813.
XX PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX Claim 1; Page; 1017pp; English.
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat DNA (shown in Table 2 of the
CC specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4789 BP; 1333 A; 981 C; 1050 G; 1425 T; 0 U; 0 Other;
Query Match 74.8%; Score 17.2; DB 9; Length 4789;
Best Local Similarity 86.4%; Pred. No. 5.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AATGAACACTACATACACCACC 23
|||||
Db 2746 AATGACACATGCCACCACC 2725
RESULT 11
ABK31277/c
ID ABK31277 standard; DNA; 9087 BP.
XX

AC ABK31277;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Signal transduction associated gene modified complementary DNA #60.
 XX
 KW Human; signal transduction associated gene; cytosine methylation state;
 KW CpG island; signal transduction associated disease; solid tumour; cancer;
 KW antitumour; cytostatic; mutant; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 OS
 FN WO200200926-A2.
 FN
 PD 03-JAN-2002.
 PD
 XX
 XX 29-JUN-2001; 2001WO-EP007472.
 PF
 XX 30-JUN-2000; 2000DE-01032529.
 XX
 PR 01-SEP-2000; 2000DE-01043826.
 PR
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2002-147896/19.
 XX
 DR Oligonucleotide for diagnosis and therapy of diseases associated with
 PT signal transduction e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with signal transduction.
 PT
 XX Claim 1; SEQ ID NO 120; 24pp; English.
 PS
 XX The present invention relates to chemically modified DNA sequences of
 CC signal transduction associated genes. The DNA sequences are chemically
 CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
 CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
 CC the cytosine methylation state (CpG islands) of these genes, and a method
 CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
 CC genes associated with signal transduction. The genomic DNA can be
 CC obtained from cells or cellular components which contain DNA, e.g. cell
 CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
 CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
 CC brain, heart, prostate, lung, breast or liver, histologic object slides,
 CC and all their possible combinations. The sequences of the invention are
 CC useful for the diagnosis and therapy of diseases associated with signal
 CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
 CC chemically pretreated genomic DNA sequences of different genes associated
 CC with signal transduction, or their complementary sequences. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office
 XX
 SQ Sequence 9087 BP; 2207 A; 253 C; 2592 G; 4034 T; 0 U; 1 Other;
 Query Match 74.8%; Score 17.2; DB 6; Length 9087;
 Best Local Similarity 86.4%; Pred. No. 5.8e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 AATGAACCTACATCAACACC 23
 |||||
 Db 2025 AACGAACCTACATCAACACC 2004
 RESULT 12
 ABL70238/c
 ID ABL70238 standard; DNA; 9087 BP.
 XX
 AC ABL70238;
 XX
 DT 01-JUL-2002 (first entry)
 XX

DE Chemically treated cell signalling DNA sequence complementary to#64.
 XX
 KW Cell signalling; cytosine methylation; cell signalling disease; cancer;
 KW tumour; cytostatic; ds.
 XX
 OS Unidentified.
 OS
 FN WO200202807-A2.
 FN
 PD 10-JAN-2002.
 PD
 XX 29-JUN-2001; 2001WO-EP007471.
 PF
 XX 30-JUN-2000; 2000DE-01032529.
 XX
 PR 01-SEP-2000; 2000DE-01043826.
 PR
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2002-154758/20.
 XX
 DR Nucleic acid, useful for diagnosis and therapy of diseases associated
 PT with cell signalling e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with cell signalling.
 PT
 XX Claim 1; SEQ ID NO 128; 24pp + Sequence Listing; English.
 PS
 XX The invention relates to a nucleic acid comprising a sequence of at least
 CC 18 bases of a segment of chemically pretreated DNA of genes associated
 CC with cell signalling. The activity of the modified sequences of the
 CC invention may be described as cytostatic. The object of the invention is
 CC to provide the chemically modified DNA of genes associated with cell
 CC signalling, as well as oligonucleotides and/or PNA-oligomers for
 CC detecting cytosine methylations, as well as a method which is
 CC particularly suitable for the diagnosis and/or therapy of genetic and
 CC epigenetic parameters of genes associated with cell signalling. The
 CC chemically modified DNA provided by the invention is useful for diagnosis
 CC and therapy of diseases such as solid tumours and cancer. The sequences
 CC given in records ABL70111-ABL70626 represent chemically pre-treated
 CC genomic DNA's of genes associated with cell signalling. Note: The
 CC sequence data for this patent is not represented in the printed
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office
 XX
 SQ Sequence 9087 BP; 2207 A; 253 C; 2592 G; 4034 T; 0 U; 1 Other;
 Query Match 74.8%; Score 17.2; DB 6; Length 9087;
 Best Local Similarity 86.4%; Pred. No. 5.8e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 AATGAACCTACATCAACACC 23
 |||||
 Db 2025 AACGAACCTACATCAACACC 2004
 RESULT 13
 AAS61181/c
 ID AAS61181 standard; DNA; 9087 BP.
 XX
 AC AAS61181;
 XX
 DT 29-JAN-2002 (first entry)
 DT
 XX Human gene regulation-associated gene oligonucleotide #136.
 DE
 XX Human; Gene regulation-associated gene; severe combined immunodeficiency;
 KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
 KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
 KW renal disease; Preeclampsia; cardiac allograft vascular disease;
 KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
 KW immunostimulant; cardiac; antiinflammatory; coagulant; antiasthmatic;
 KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.

```

XX OS Homo sapiens.
XX PN WO200177375-A2.
XX PD 18-OCT-2001.
XX PF 06-APR-2001; 2001WO-EP003368.
XX PR 06-APR-2000; 2000DE-01019058.
XX PR 07-APR-2000; 2000DE-01019173.
XX PR 30-JUN-2000; 2000DE-01032529.
XX PR 01-SEP-2000; 2000DE-01043626.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-017470/02.
XX PT New nucleic acid sequences from chemically modified genes associated with
XX gene regulation, useful for analyzing cytosine methylations for diagnosis
XX and therapy of diseases e.g. severe combined immunodeficiency disease.
XX PS Claim 1; SEQ ID NO 140; 26pp; English.
XX CC The invention relates to 224 nucleic acid sequences comprising at least
XX 18 bases of a chemically pretreated gene associated with gene regulation
XX selected from 43 known genes (or complementary sequences). The chemical
XX pretreatment converts cytosine bases unmethylated at the 5-position to
XX uracil or another base with hybridisation behaviour dissimilar to
XX cytosine, to enable analysis of cytosine methylations. The DNA sequences,
XX oligomers (or sets/arrays) and method are useful in the diagnosis of
XX diseases (or predisposition to diseases) associated with gene regulation
XX and in therapy of such diseases, by enabling analysis of the cytosine
XX methylation patterns of such genes, kits are provided. They are
XX especially useful in diagnosis and therapy of e.g. severe combined
XX immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
XX and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
XX syndrome, renal disease, preclampsia, graft versus-host disease. The
XX present sequence is a sequence included in the sequence data for this
XX specification and is associated with the human gene regulation-associated
XX genes. Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO.int/pub/pub/published_pct_sequences
XX SQ Sequence 9087 BP; 2207 A; 253 C; 2592 G; 4034 T; 0 U; 1 Other;

Query Match 74.8%; Score 17.2; DB 6; Length 9087;
Best Local Similarity 86.4%; Pred. No. 5.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AATGAACCTACATACAAACACC 23
Db 2025 AACGAACCTACATACAAACACC 2004

RESULT 14
ABK92230/c
ID ABK92230 standard; DNA; 12879 BP.
XX AC ABK92230;
XX DT 15-AUG-2002 (first entry)
XX DE Prostate cancer-associated DNA sequence #116.
XX KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;
XX gene therapy; gene; ds.
XX OS Mammalia.
XX PN WO200230268-A2.

```

```

XX PD 18-APR-2002.
XX PF 12-OCT-2001; 2001WO-US032045.
XX PR 13-OCT-2000; 2000US-00687576.
XX PR 08-DEC-2000; 2000US-00733288.
XX PR 08-DEC-2000; 2000US-00733742.
XX PR 24-JAN-2001; 2001US-0263957P.
XX PR 16-MAR-2001; 2001US-0276791P.
XX PR 16-MAR-2001; 2001US-0276888P.
XX PR 06-APR-2001; 2001US-0281922P.
XX PR 24-APR-2001; 2001US-0286214P.
XX PR 30-APR-2001; 2001US-00847046.
XX PR 04-MAY-2001; 2001US-0288589P.
XX PA (EOSB-) EOS BIOTECHNOLOGY INC.
XX PI Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;
XX DR WPI; 2002-471335/50.
XX DR P-PSDB; ABG61913.
XX PT Detecting a prostate cancer-associated transcript in a cell in a patient,
XX useful for diagnosing prostate cancer (PC) or screening modulators of PC,
XX by determining if prostate cancer-associated genes are expressed in a
XX prostate tissue.
XX PS Claim 22; Page 394-397; 436pp; English.
XX CC The present invention relates to methods of detecting a prostate cancer-
XX associated transcript in a cell from a patient. The method comprises
XX contacting a biological sample from the patient with prostate cancer-
XX associated polynucleotides (designated PC genes) that selectively
XX hybridise to a sequence that is at least 80% identical to them. The
XX prostate cancer-associated polynucleotide sequences are differentially
XX expressed in prostate tumour tissue or in prostate cancer and are derived
XX from the tissues of various organisms such as humans or other mammals
XX (e.g. mice, sheep and dogs). The methods of the invention are useful for
XX diagnosing and treating prostate cancer in mammals. The prostate cancer-
XX associated genes are useful for diagnosing or treating prostate cancer,
XX as well as for identifying modulators of prostate cancer or agents that
XX inhibit prostate cancer. The nucleic acid sequences are particularly
XX useful in gene therapy, as a vaccine or in antisense applications.
XX CC ABK92115-ABK92263 represent prostate cancer-associated polynucleotide
XX sequences
XX SQ Sequence 12879 BP; 3852 A; 2674 C; 2768 G; 3585 T; 0 U; 0 Other;

Query Match 74.8%; Score 17.2; DB 6; Length 12879;
Best Local Similarity 86.4%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AATGAACCTACATACAAACACC 23
Db 11765 AATGAACCTACATACATCACC 11744

RESULT 15
AAK51828/c
ID AAK51828 standard; cDNA; 13202 BP.
XX AC AAK51828;
XX DT 06-NOV-2001 (first entry)
XX DE Human polynucleotide SEQ ID NO 373.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation; ss.

```

OS Homo sapiens.

XX WO200157190-A2.

XX

XX 09-AUG-2001.

XX

PF 05-FEB-2001; 2001WO-US004098.

XX

XX 03-FEB-2000; 2000US-00496914.

PR 27-APR-2000; 2000US-00560875.

PR 20-JUN-2000; 2000US-00598075.

PR 19-JUL-2000; 2000US-00620325.

PR 01-SEP-2000; 2000US-00654936.

PR 15-SEP-2000; 2000US-00663561.

PR 20-OCT-2000; 2000US-00693325.

PR 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.

XX

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI: 2001-476283/51.

DR P-PSDB; AAM78695.

XX

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful

PT in diagnosis and gene therapy.

XX

XX Claim 1; Page 1414-1426; 6221pp; English.

PS

PS The invention relates to polynucleotides (AAK51456-AAK53435) and the

CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111

CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the

CC sequence listing were missing at the time of publication

XX

SQ Sequence 13202 BP; 3921 A; 2764 C; 2865 G; 3652 T; 0 U; 0 Other;

Query Match 74.8%; Score 17.2; DB 4; Length 13202;

Best Local Similarity 86.4%; Pred.No. 5.9e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AATGAACACTACATAACCAACC 23
|||||

Db 12091 AATGAACCTCATTACATCAC 12070
|||||

Search completed: March 24, 2004, 22:28:19

Job time : 124.427 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 21:57:32 ; Search time 1119.76 Seconds
(without alignments)
613.370 Million cell updates/sec

Title: US-09-889-611A-19

Perfect score: 23
Sequence: 1 gaatgaactacatacaaccacc 23

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_estc.*

9: gb_est1.*

10: gb_est2.*

11: gb_est3.*

12: gb_est4.*

13: gb_est5.*

14: gb_est6.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	18.4	80.0	569	28	CC162058
c 2	18.4	80.0	684	28	AZ400233
c 3	18.4	80.0	782	29	CG114080
c 4	18.4	80.0	785	28	CC090422

5	18.4	80.0	927	28	CC134385
6	18.4	80.0	999	10	BE572261
c 7	18.4	80.0	1201	9	AL572872
c 8	18.2	79.1	129	10	AW698814
c 9	18.2	79.1	307	14	CD183415
c 10	18.2	79.1	309	14	CD133709
c 11	18.2	79.1	313	14	CD188719
c 12	18.2	79.1	317	14	CD165299
c 13	18.2	79.1	531	14	CD189864
c 14	18.2	79.1	537	14	CD083615
c 15	18.2	79.1	556	14	CD073220
c 16	18.2	79.1	648	14	CD074368
c 17	18.2	79.1	762	14	CF443253
c 18	18.2	79.1	763	10	BE387554
c 19	18.2	79.1	813	29	CG442765
c 20	18.2	79.1	845	29	CG442758
c 21	18.2	79.1	294	10	BF365378
c 22	17.8	77.4	72	14	CF608248
c 23	17.8	77.4	180	28	AZ442459
c 24	17.8	77.4	217	9	AI901912
c 25	17.8	77.4	247	9	AV208603
c 26	17.8	77.4	333	28	AQ908428
c 27	17.8	77.4	431	28	AQ680312
c 28	17.8	77.4	456	13	BY548796
c 29	17.8	77.4	503	28	BH193866
c 30	17.8	77.4	625	28	BZ484996
c 31	17.8	77.4	632	28	BH844430
c 32	17.8	77.4	633	28	BH509041
c 33	17.8	77.4	649	14	CF207448
c 34	17.8	77.4	652	28	AZ606165
c 35	17.8	77.4	666	28	BH247014
c 36	17.8	77.4	682	28	BH841120
c 37	17.8	77.4	694	28	BZ005096
c 38	17.8	77.4	699	28	BH655941
c 39	17.8	77.4	701	28	BH960009
c 40	17.8	77.4	708	28	AZ008960
c 41	17.8	77.4	713	28	BZ678717
c 42	17.8	77.4	766	28	BH490439
c 43	17.8	77.4	782	14	CF066855
c 44	17.8	77.4	820	28	BZ487245
c 45	17.8	77.4	890	28	BZ389277

ALIGNMENTS

RESULT 1
CC162058/c
LOCUS
DEFINITION
CC162058
WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
Genomic survey sequence.
CC162058.1 GI:30188242
GSS.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Zea mays

Eukaryota;

Viridiplantae;

Streptophyta;

Embryophyta;

Tracheophyta;

Spermatophyta;

Magnoliophyta;

Liliopsida;

Poales;

Poaceae;

PACCAD

clade; Panicoidae;

Andropogoneae; Zea.

1 (bases 1 to 569)

Rabinowicz P.D., O'Shaughnessy A.L., Ballia V., Dedhia N.,

Katzburger F., King L., Miller B., Muller S., Nascimento L.,

Zutavern T., McCombie W.R. and Martienssen R.A.

Unpublished (2002)

Unpublished (2002)

Contact: W. Richard McCombie

CC134385 NDL.60D18
BE572261 60130149
AL572872 AL572872
AW698814 r406 non-
CD183415 MS1-003BU
CD133709 MG1-0024U
CD188719 MS1-0063U
CD165299 ML1-0095T
CD189864 MS1-0063U
CD083615 MA3-9999U
CD073220 MA3-0001U
CD074368 MA3-0001U
CF443253 EST679598
BE387554 601274341
CG442765 OGVGP70TV
CG442758 OGVGP70TH
BF365378 QVA-NT002
CF608248 GRMA01.0
AZ442459 IM0236B06
AI901912 618012C05
AV208603 AV208603
AQ908428 GSSTC0518
AQ680312 HS 2152 B
BY548796 BY548796
BH193866 TC3-23D6.
BZ484996 BONA787T
BH844430 TC3-57118
BH509041 BOGRJ88TR
CF207448 CAB20001
AZ606165 IM0427L21
BH247014 BOGA068TF
BH841120 TC3-55C20
BZ005096 oe127d10.
BH655941 BOXA70TF
BH960009 od108a10.
AZ008960 RPCI-23-3
BZ678717 PUBAH86TD
BH490439 BOCON20TF
CF066855 Ac5378 Am
BZ487245 BONOO71TF
BZ389277 EINDI54TR

Seq primer: -21M13UnivFwd
 Class: shotgun
 High quality sequence stop: 569.
 Location/Qualifiers

FEATURES

1..569
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone="ii181b04"
 /lab_host="DH5a"
 /clone_lib="WGS-ZmaysF (DH5a methyl filtered)"
 /note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;
 The vector was digested with XbaI and one nucleotide was
 added by fill in the recessive 3' end. The genomic DNA
 was nebulized, end repaired, adaptor ligated and size
 fractionated using sephadex. The resulting fragments were
 between 0.8 and 3 kb and were cloned into the vector (.x/y
 reads in M13mp19, .b/g reads in pUC19). The same ligation
 was transformed into DH5a."

ORIGIN

Query Match 80.0%; Score 18.4; DB 28; Length 569;
 Best Local Similarity 95.0%; Pred. No. 2.6e+03;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 4 TGAACACTACATCAACACCACC 23
 ||||||||||||||||||
 Db 59 TGGACTACATCAACACCACC 40

RESULT 2

AZ400233
 LOCUS AZ400233 684 bp DNA linear GSS 03-OCT-2000
 DEFINITION LM0165121F Mouse 10kb plasmid UUC1M library Mus musculus genomic
 clone UUC1M0165121 F, genomic survey sequence.

ACCESSION

VERSION AZ400233.1 GI:10515307

KEYWORDS

GSS.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 684)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Ismail, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,
 Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von
 Niederhausern, A. and Wright, D., Weiss, R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL

COMMENT

Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0166 Row: L Column: 21

Seq primer: CTTTGAACACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 684.

FEATURES

Location/Qualifiers

1..684

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUC1M0165121"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"

/clone_lib="Mouse 10kb plasmid UUC1M library"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (GI4732114|GB|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

ORIGIN

Query Match 80.0%; Score 18.4; DB 28; Length 684;
 Best Local Similarity 95.0%; Pred. No. 2.6e+03;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 2 AATGAACACTACATCAACACCA 21
 ||||||||||||||||||
 Db 393 AATGAACACTACATCAACACCA 412

RESULT 3

CG114080/c
 LOCUS CG114080 782 bp DNA linear GSS 20-AUG-2003
 DEFINITION PUJGG81TD ZM_0.6.1.0 KB Zea mays genomic clone ZM8Bta0661N17,
 genomic survey sequence.

ACCESSION

VERSION CG114080.1 GI:33997517

KEYWORDS

GSS.

SOURCE

Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 782)
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
 Reenick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
 Bennetzen, J.

TITLE

Maize Genomics Consortium

JOURNAL

COMMENT

Unpublished (2003)
 Other GSSs: PUJGG81TB
 Contact: Cathy Whitelaw
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: TF

Class: sheared ends.

FEATURES

Location/Qualifiers

1..782

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZM8Bta0661N17"

/clone_lib="ZM_0.6.1.0 KB"

/note="vector: PCR4-TOFO; Site 1: EcoRI; 0.6-1.0 kb high
 Cot selected genomic DNA library"

ORIGIN

Query Match 80.0%; Score 18.4; DB 29; Length 782;
 Best Local Similarity 95.0%; Pred. No. 2.6e+03;

```

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 TGAAGTACATACACACACC 23
    |||||
Db 500 TGGACTACATACACACACC 481

CC090422 785 bp DNA linear GSS 16-APR-2003
CSU-K33r.9P3.T7 CSU-K33r Aedes aegypti genomic clone CSU-K33r.9P3,
genomic survey sequence.
CC090422
CC090422.1 GI:29947659
GSS.
SOURCE Aedes aegypti (yellow fever mosquito)
ORGANISM Aedes aegypti
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
Stegomyia.
REFERENCE 1 (bases 1 to 785)
AUTHORS Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
TITLE End sequencing of Aedes aegypti BACs
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: CSU-K33r.9P3.SP6
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by Susan Brown and Dennis Knudson at Colorado
State University.
Seq primer: T7
Class: BAC ends.

FEATURES             Location/Qualifiers
     source
     1..785
        /organism="Aedes aegypti"
        /mol_type="genomic DNA"
        /strain="Rexville"
        /db_xref="taxon:7159"
        /clone="CSU-K33r.9P3"
        /clone_lib="CSU-K33r"
        /notes="Vector: pBelOBAC11; Site_1: HindIII"

ORIGIN
Query Match      80.0%; Score 18.4; DB 28; Length 785;
Best Local Similarity 95.0%; Pred. No. 2.6e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAATGAAGTACATACACACC 20
    |||||
Db 380 GAATGAAGTACATACACACC 361

CC134385 927 bp DNA linear GSS 16-APR-2003
NDL.60D18.SP6 Notre Dame Liverpool Aedes aegypti genomic clone
NDL.60D18, genomic survey sequence.
CC134385
CC134385.1 GI:30003440
GSS.
SOURCE Aedes aegypti (yellow fever mosquito)
ORGANISM Aedes aegypti
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
Stegomyia.
REFERENCE 1 (bases 1 to 927)
AUTHORS Loftus,B., Shetty,J., Knudson,D. and Severson,D.
TITLE BAC end sequencing of Aedes aegypti
JOURNAL Unpublished (2003)

```

```

Other_GSSs: NDL.60D18.T7
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by David Severson
Seq primer: SP6
Class: BAC ends.

FEATURES             Location/Qualifiers
     source
     1..927
        /organism="Aedes aegypti"
        /mol_type="genomic DNA"
        /strain="liverpool"
        /db_xref="taxon:7159"
        /clone="NDL.60D18"
        /clone_lib="Notre Dame Liverpool"
        /notes="Vector: pECBAC1; Site_1: Hind III; The library was
        prepared from whole body tissue of newly hatched L1 larvae
        by David Severson at the University of Notre Dame and
        Hongbin Zhang"

ORIGIN
Query Match      80.0%; Score 18.4; DB 28; Length 927;
Best Local Similarity 95.0%; Pred. No. 2.6e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAATGAAGTACATACACACC 20
    |||||
Db 577 GAATGAAGTACATACACACC 596

RESULT 6
BE572261 999 bp mRNA linear EST 15-AUG-2000
601330149F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3707436 5',
mRNA sequence.
ACCESSION BE572261
VERSION BE572261.1 GI:9815981
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 999)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM8925 row: b column: 13
High quality sequence stop: 340.

FEATURES             Location/Qualifiers
     source
     1..999
        /organism="Mus musculus"
        /mol_type="mRNA"
        /strain="FVB/N"
        /db_xref="taxon:10090"
        /clone="IMAGE:3707436"
        /sex="female, virgin"
        /tissue type="infiltrating ductal carcinoma"
        /dev stage="5 months"
        /lab_host="DH10B"
        /clone_lib="NCI_CGAP_Mam6"

```

/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI; Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

ORIGIN

Query Match 80.0%; Score 18.4; DB 10; Length 999;
Best Local Similarity 95.0%; Pred. No. 2.6e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TGAACACTACATAACACACC 23
||||| : |||||
Db 464 TGAACCACTACATACACACC 483

RESULT 7

AL572872/c 1201 bp mRNA linear EST 31-MAY-2003
LOCUS
DEFINITION AL572872 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSODI034YH01 3-PRIME, mRNA sequence.

ACCESSION
VERSION AL572872
KEYWORDS EST.

SOURCE
Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Li, W.B., Gruber, C., Jesses, J., and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT On Feb 16, 2001 this sequence version replaced gi:12931557.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 2214.r For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CSODI034CD01NP1&cluster=2214.r. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CSODI034CD01NP1.

FEATURES

source

1. 1201

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSODI034YH01"

/tissue_type="PLACENTA COT 25-NORMALIZED"

/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo (dt)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 80.0%; Score 18.4; DB 9; Length 1201;
Best Local Similarity 86.4%; Pred. No. 2.6e+03;
Matches 19; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAATGAACACTACATAACACACC 22
||||| : |||||
Db 526 GAATGAATAAATAACACACC 505

RESULT 8

AW698814/c 129 bp mRNA linear EST 17-APR-2000
LOCUS
DEFINITION AW698814 r406 non-glandular-haired subtracted cDNA library Medicago sativa
cDNA, mRNA sequence.

ACCESSION AW698814

VERSION
KEYWORDS
SOURCE
ORGANISM

Medicago sativa

Medicago sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids 1; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

1 (bases 1 to 129)

Hays, D. and Skinner, D.

Expressed sequence tags subtracted in reciprocal fashion between glandular-haired and non-glandular-haired full sibs of alfalfa unpublished (2000)

JOURNAL

COMMENT

Contact: Hays DB

USDA, ARS, GMPRC, PSERU; Department of Agronomy

Kansas State University

Throckmorton Hall, Manhattan, KS 66506, USA

Tel: 785 532 7116

Fax: 785 532 6167

Email: dhays@genes.alfalfa.ksu.edu

Seq primer: SP6.

FEATURES

source

1. 129

/organism="Medicago sativa"

/mol_type="mRNA"

/cultivar="Riley X KS224"

/db_xref="taxon:3879"

/tissue_type="leaf and stem"

/clone_lib="non-glandular-haired subtracted cDNA library"

/note="glandular-haired versus non-glandular-haired reciprocal cDNA subtraction with CLONTECH PCR-Select cDNA subtraction."

ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 129;
Best Local Similarity 87.0%; Pred. No. 3.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAATGAACACTACATAACACACC 23
||||| : |||||
Db 88 GAATGAACATAAATAACACACC 66

RESULT 9

LOCUS

DEFINITION

CD183415/c

MS1-00380-A246-G09-U.B

MS1-00380-A246-G09.B, mRNA sequence.

ACCESSION

CD183415

VERSION

CD183415.1

KEYWORDS

SOURCE

ORGANISM

Schistosoma mansoni

Schistosoma mansoni

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigeida; Schistosomatoidea; Schistosomatidae; Schistosoma.

1 (bases 1 to 307)

AUTHORS

Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,

Opopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr.,

Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,

Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,

Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,

Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,

Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,

Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M.,

Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.

Transcriptome analysis of the acclomate human parasite Schistosoma

mansoni

JOURNAL

MEDLINE

COMMENT

Nat. Genet. 35 (2), 148-157 (2003)

Contact: Dr. Sergio Verjovski-Almeida

Departamento de Bioquímica

Instituto de Química - Universidade de São Paulo

Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,

Brasil

Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjoeiq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST Genome Project. All sequences in the project were assembled and annotated. This entry and all the assembled sequences can be seen in the following URL <http://bioinfo.iq.usp.br/schisto/>
Plate: MS1-0038U-A246 row: 9 column: G.

FEATURES

source
1..307

/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MS1-0038U-A246-G09.B"
/sex="mixed pool"
/dev_stage="schistosomulum"
/lab_host="in vitro culture"
/clone_lib="MS1-0038"
/note="Vector: pGEM T-easy"

ORIGIN

Query Match 79.1%; Score 18.2; DB 14; Length 307;

Best Local Similarity 87.0%; Pred. No. 3.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAATGAACCTACATAACACACC 23

|||||
Db 128 GAATAAATCTACATAACACACC 106

RESULT 10

CD133709

LOCUS

CD133709 309 bp mRNA linear EST 14-SEP-2003
MS1-0024U-A317-H10-U.G. MS1-0024 Schistosoma mansoni CDNA clone

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatidae; Schistosomatidae; Schistosoma.

1 (bases 1 to 309)
Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M., Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr., Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F., Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L., Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A., Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A., Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T., Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M., Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.
Transcriptome analysis of the acelomate human parasite Schistosoma mansoni

TITLE

Nat. Genet. 35 (2), 148-157 (2003)

JOURNAL

MEDLINE

COMMENT

Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP, Brasil

Tel: +55-11-3091-2173

Fax: +55-11-3091-2186

Email: verjoeiq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST Genome Project. All sequences in the project were assembled and annotated. This entry and all the assembled sequences can be seen in the following URL <http://bioinfo.iq.usp.br/schisto/>
Plate: MG1-0024U-A317 row: 10 column: H.

FEATURES

source
1..309

/organism="Schistosoma mansoni"

/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MG1-0024U-A317-H10.G"
/sex="mixed pool"
/dev_stage="germball"
/lab_host="Biomphalaria glabrata"
/clone_lib="MG1-0024"
/note="Vector: pGEM T-easy"

ORIGIN

Query Match 79.1%; Score 18.2; DB 14; Length 309;

Best Local Similarity 87.0%; Pred. No. 3.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GAATGAACCTACATAACACACC 23

|||||
Db 180 GAATAAATCTACATAACACACC 202

RESULT 11

CD188719/c

LOCUS

CD188719 313 bp mRNA linear EST 14-SEP-2003
MS1-0063U-A262-A04-U.B MS1-0063 Schistosoma mansoni CDNA clone

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 313)
Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M., Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.Y. Jr., Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F., Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L., Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A., Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A., Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T., Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.M., Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.
Transcriptome analysis of the acelomate human parasite Schistosoma mansoni

TITLE

JOURNAL

MEDLINE

COMMENT

Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquímica
Instituto de Química - Universidade de São Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP, Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjoeiq.usp.br
This sequence was derived from the FAPESP Schistosoma mansoni EST Genome Project. All sequences in the project were assembled and annotated. This entry and all the assembled sequences can be seen in the following URL <http://bioinfo.iq.usp.br/schisto/>
Plate: MS1-0063U-A262 row: 4 column: A.

FEATURES

source
1..313

/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="MS1-0063U-A262-A04.B"
/sex="mixed pool"
/dev_stage="schistosomulum"
/lab_host="in vitro culture"
/clone_lib="MS1-0063"
/note="Vector: pGEM T-easy"

ORIGIN

Query Match 79.1%; Score 18.2; DB 14; Length 313;

Best Local Similarity 87.0%; Pred. No. 3.1e+03;

Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
Transcriptome analysis of the acelomate human parasite Schistosoma
mansoni

Nat. Genet. 35 (2), 148-157 (2003)

22879926

Contact: Dr. Sergio Verjovski-Almeida

Departamento de Bioquímica

Instituto de Química - Universidade de São Paulo

Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil

Tel: +55-11-3091-2173

Fax: +55-11-3091-2186

Email: verjo@iq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL: <http://bioinfo.iq.usp.br/schisto/>
Plate: MA3-9999U-V243 row: 11 column: H.

FEATURES

source

```
1..537
    Location/Qualifiers
        /organism="Schistosoma mansoni"
        /mol_type="mRNA"
        /db_xref="taxon:6183"
        /clone="MA3-9999U-V243-H11.B"
        /sex="mixed pool"
        /dev_stage="adult"
        /lab_host="Mus musculus"
        /clone_lib="MA3-0001"
```

ORIGIN

Query Match 79.1%; Score 18.2; DB 14; Length 537;
Best Local Similarity 87.0%; Pred. No. 3.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 GAATGAACACTACATACACACC 23
   |||||
DB 460 GAATAAACTACATACACACC 438
```

RESULT 15

CD073220

LOCUS

MA3-0001U-L241-H07-U-B MA3-0001 Schistosoma mansoni cDNA clone

MA3-0001U-L241-H07.B, mRNA sequence.

CD073220

ACCESSION

CD073220.1 GI:34624266

KEYWORDS

EST.

SOURCE

Schistosoma mansoni

Schistosoma mansoni

Bukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigeidida; Schistosomatidae; Schistosomatidae; Schistosoma.

1 (bases 1 to 556)

Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M.,
Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr.,
Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F.,
Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L.,
Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A.,
Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A.,
Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T.,
Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M.,
Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
Transcriptome analysis of the acelomate human parasite Schistosoma
mansoni

Nat. Genet. 35 (2), 148-157 (2003)

22879926

Contact: Dr. Sergio Verjovski-Almeida

Departamento de Bioquímica

Instituto de Química - Universidade de São Paulo

Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
Brasil

TITLE

JOURNAL

MEDLINE

COMMENT

Tel: +55-11-3091-2173

Fax: +55-11-3091-2186

Email: verjo@iq.usp.br

This sequence was derived from the FAPESP Schistosoma mansoni EST
Genome Project. All sequences in the project were assembled and
annotated. This entry and all the assembled sequences can be seen
in the following URL: <http://bioinfo.iq.usp.br/schisto/>
Plate: MA3-0001U-L241 row: 7 column: H.

FEATURES

source

```
1..556
    Location/Qualifiers
        /organism="Schistosoma mansoni"
        /mol_type="mRNA"
        /db_xref="taxon:6183"
        /clone="MA3-0001U-L241-H07.B"
        /sex="mixed pool"
        /dev_stage="adult"
        /lab_host="Mus musculus"
        /clone_lib="MA3-0001"
```

ORIGIN

Query Match 79.1%; Score 18.2; DB 14; Length 556;
Best Local Similarity 87.0%; Pred. No. 3.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 GAATGAACACTACATACACACC 23
   |||||
DB 197 GAATAAACTACATACACACC 219
```

Search completed: March 25, 2004, 00:13:45
Job time : 1125.76 secs


```

REFERENCE
AUTHORS
TITLE
JOURNAL
SUBMITTED (15-FEB-2000) Institute of Medical Sciences and
Department of Internal Medicine, Tokai University School of
Medicine, Bohseidai, Isehara, Kanagawa 259-1193, Japan
3 (bases 1 to 4229)
Miyata, T., Nangaku, M., Inagi, R. and Kurokawa, K.
Direct Submission
TITLE
JOURNAL
SUBMITTED (29-MAR-2002) Institute of Medical Sciences and
Department of Internal Medicine, Tokai University School of
Medicine, Bohseidai, Isehara, Kanagawa 259-1193, Japan
Sequence update by submitter
On Mar 29, 2002 this sequence version replaced gi:18000453.
FEATURES
source
1. 4229
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21.3"
1. 4021
promoter
4022. >4229
mRNA
/product="megsin"
ORIGIN
Query Match 93.6%; Score 23.4; DB 9; Length 4229;
Best Local Similarity 96.0%; Pred. No. 1.7;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AACACCTTAGTCAGACTACTT 25
Db 3936 AACACCTTAGTCAGACTACTT 3960
RESULT 2
AC072051 157284 bp DNA linear PRI 03-DEC-2001
LOCUS Homo sapiens chromosome, clone RP11-79D21, complete sequence.
AC072051
AC072051.8 GI:16974280
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157284)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome, clone RP11-79D21
Unpublished
2 (bases 1 to 157284)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bede, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castie, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, J., Oliviar, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
REFERENCE
AUTHORS
TITLE
JOURNAL
SUBMITTED (07-JUN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 157284)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fargo, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gird, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
MacLear, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
SUBMITTED (25-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 157284)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fargo, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gird, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
MacLear, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
SUBMITTED (03-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 18, 2001 this sequence version replaced gi:14277307.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L966
Center clone name: 79_D_21
-----
Location/Qualifiers
1. 157284
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-79D21"

```


Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menes, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roberti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 1, 2002 this sequence version replaced gi:11990731.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L8516
Center clone name: 317_G_1

* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 85659: contig of 85659 bp in length
* 85660 85759: gap of 100 bp
* 85760 163014: contig of 77255 bp in length
* 163015 163114: gap of 100 bp
* 163115 189092: contig of 25978 bp in length.

FEATURES
source

1. 189092
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-317G1"
/clone_lib="RP11-11 Human Male BAC"

ORIGIN

Query Match 93.6%; Score 23.4; DB 2; Length 189092;
Best Local Similarity 96.0%; Pred. No. 1;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACACTACTTT 25
|||||

Db 66579 AACCACTTAGTCAGATACACTACTTT 66603
|||||

RESULT 4
AX346917/C

LOCUS AX346917 6182 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 1988 from Patent WO0200928.
ACCESSION AX346917

AX346917.1 GI:18494803

KEYWORDS
SOURCE

ORGANISM
synthetic construct
artificial sequences.

REFERENCE

1 Olek, A., Piepenbrock, C. and Berlin, K.
Diagnosis of diseases associated with the immune system
Patent: WO 0200928-A 1988 03-JAN-2002;
Epigenomics AG (DE)

FEATURES

source
Location/Qualifiers
1. 6182
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 80.8%; Score 20.2; DB 6; Length 6182;
Best Local Similarity 88.0%; Pred. No. 58;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACACTACTTT 25
|||||

Db 1295 AACCACTTAGTCAGATACACTACTTT 1271
|||||

RESULT 5

LOCUS

AC111981 346601 bp DNA linear HTG 19-NOV-2002
Rattus norvegicus clone CH210-121G8, WORKING DRAFT SEQUENCE, 5
unordered pieces.

ACCESSION

AC111981.4 GI:25073004

HTG: HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

KEYWORDS

Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 346601)
Muzny, D., Marie, Metzker, M., Lee, A., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Fallis, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmood, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwaokemele, O., Okwunonu, G., Olanpunsagoon, A., Pal, S., Parks, K.,

RP11-12C17 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECIOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-12C17. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP11-12C17 is at 63011 in this sequence. The true left end of clone GSI-11815 is at 33117 in this sequence. The true right end of clone RP11-6713 is at 100 in this sequence.

FEATURES

```

source
1..63011
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="1"
  /clone="RP11-12C17"
  /clone_lib="RPCI-11.1"
repeat_region
1..238
  /note="THE1C repeat: matches 1..243 of consensus"
repeat_region
312..598
  /note="HERV23 repeat: matches 1..274 of consensus"
  /complement(576..1195)
misc_feature
609..839
  /note="match: GSS: Em:AQ530043"
repeat_region
1093..1195
  /note="LTR24 repeat: matches 250..480 of consensus"
repeat_region
1796..1913
  /note="LTR24 repeat: matches 1..109 of consensus"
repeat_region
2084..2295
  /note="L2 repeat: matches 2617..2735 of consensus"
repeat_region
2368..2635
  /note="TIGER1 repeat: matches 863..1014 of consensus"
repeat_region
2636..3044
  /note="MER44C repeat: matches 81..359 of consensus"
repeat_region
3045..3405
  /note="L2 repeat: matches 2069..2441 of consensus"
repeat_region
3407..4975
  /note="THE1B repeat: matches 1..364 of consensus"
repeat_region
4976..5343
  /note="THE1B-INTERNAL repeat: matches 2..1580 of consensus"
repeat_region
5344..5423
  /note="THE1B repeat: matches 1..364 of consensus"
repeat_region
5424..5743
  /note="L2 repeat: matches 2441..2513 of consensus"
repeat_region
5744..5847
  /note="AluYb repeat: matches 1..299 of consensus"
repeat_region
7238..7379
  /note="L2 repeat: matches 2513..2722 of consensus"
repeat_region
7693..7982
  /note="MIR repeat: matches 40..182 of consensus"
repeat_region
8098..8156
  /note="AluSp repeat: matches 1..312 of consensus"
repeat_region
8157..8456
  /note="MIR repeat: matches 121..182 of consensus"
repeat_region
8457..8574
  /note="AluX repeat: matches 1..297 of consensus"
repeat_region
8804..9114
  /note="MIR repeat: matches 16..121 of consensus"
repeat_region
9593..9897
  /note="AluY repeat: matches 1..306 of consensus"
repeat_region
9898..10253
  /note="AluY repeat: matches 1..304 of consensus"
repeat_region
10470..10599
  /note="THE1B repeat: matches 1..364 of consensus"
repeat_region
10610..10903
  /note="FLAN C repeat: matches 1..132 of consensus"
repeat_region
10982..11180
  /note="AluSg repeat: matches 1..293 of consensus"
repeat_region
12003..12029
  /note="L2 repeat: matches 1986..2204 of consensus"
  /note="MIR repeat: matches 79..102 of consensus"
  /note="LTR8 repeat: matches 1..691 of consensus"
  /note="MIR repeat: matches 102..205 of consensus"
  /note="MER41A repeat: matches 1..554 of consensus"
  /note="MER41-internal repeat: matches 1..86 of consensus"
  /complement(14302..14756)
  /note="match: STS: Em:HSP03H12"
repeat_region
14302..14528
  /note="MIR repeat: matches 3..259 of consensus"
repeat_region
14622..14743
  /note="MIR repeat: matches 85..212 of consensus"
repeat_region
15285..15543
  /note="L2 repeat: matches 2255..2518 of consensus"
repeat_region
15846..16227
  /note="L1M4 repeat: matches 3865..4250 of consensus"
repeat_region
16251..16440
  /note="MER20 repeat: matches 1..179 of consensus"
repeat_region
16446..16803
  /note="MER51A repeat: matches 1..371 of consensus"
repeat_region
16804..18061
  /note="MER51-internal repeat: matches 1..1250 of consensus"
  /complement(18014..18332)
  /note="match: GSS: Em:AQ254526"
  /complement(18141..18319)
  /note="match: GSS: Em:AQ023866"
  /note="match: GSS: Em:AZ707492"
  /complement(18144..18333)
  /note="match: GSS: Em:AZ468762"
  /note="22 copies 2 mer at 75% conserved"
  /complement(18154..18330)
  /note="match: GSS: Em:AQ417380"
  /complement(18178..18350)
  /note="match: GSS: Em:AZ016479"
  /complement(18181..18319)
  /note="match: GSS: Em:B40172"
  /note="match: GSS: Em:AQ683639"
  /complement(18187..18302)
  /note="match: GSS: Em:AQ106777"
  /note="10 copies 4 mer at 85% conserved"
  /complement(18306..18333)
  /note="14 copies 2 mer at 96% conserved"
  /note="LTR10C repeat: matches 188..580 of consensus"
  /note="trNA-Lys-AAG repeat: matches 1..76 of consensus"
  /complement(18830..18860)
  /note="LTR10C repeat: matches 580..609 of consensus"
  /complement(18861..18926)
  /note="MER51-internal repeat: matches 1186..1250 of consensus"
  /note="MER51-internal repeat: matches 1181..1593 of consensus"
  /note="MER42 repeat: matches 1..356 of consensus"
  /complement(19630..20581)
  /note="MER51-internal repeat: matches 1593..2802 of consensus"
  /complement(20582..20890)
  /note="AluSg1 repeat: matches 1..309 of consensus"
  /complement(20891..21410)
  /note="MER51-internal repeat: matches 2802..3327 of

```

```

consensus"
repeat_region 21393..21893
/note="MER57-internal repeat: matches 3520..4039 of
consensus"
repeat_region 21873..21983
/note="MER61-internal repeat: matches 1584..1699 of
consensus"
repeat_region 21937..24469
/note="MER51-internal repeat: matches 5199..7816 of
consensus"
repeat_region 24471..24809
/note="MER51A repeat: matches 1..371 of consensus"
repeat_region 24852..25138
/note="AluJo repeat: matches 5..294 of consensus"
repeat_region 25144..25215
/note="LIM4 repeat: matches 3768..3836 of consensus"

Query Match 79.2%; Score 19.8; DB 9; Length 63011;
Best Local Similarity 91.3%; Pred. No. 66;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ACCACCTTACTGACATGCTACTT 24
Db 31590 ACCACCTTACTGACATGCTACTT 31568

RESULT 7
AC009946/c
LOCUS
DEFINITION Homo sapiens clone RP11-12C17, complete sequence.
ACCESSION AC009946
VERSION AC009946.2 GI:6604542
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 169072)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE The sequence of Homo sapiens clone
JOURNAL Waterston,R.H.
REFERENCE 2 (bases 1 to 169072)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 169072)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 169072)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Dec 20, 1999 this sequence version replaced gi:5836211.
FEATURES
source
location/Qualifiers
1..169072
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-12C17"

ORIGIN
Query Match 79.2%; Score 19.8; DB 9; Length 169072;
Best Local Similarity 91.3%; Pred. No. 58;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 ACCACCTTACTGACATGCTACTT 24
Db 137651 ACCACCTTACTGACATGCTACTT 137629

RESULT 8
AC139347
LOCUS
DEFINITION Mus musculus chromosome UNK clone RP24-334F11, WORKING DRAFT
SEQUENCE, 10 unordered pieces.
ACCESSION AC139347
VERSION AC139347.2 GI:28191622
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 205202)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 205202)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 205202)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Feb 1, 2003 this sequence version replaced gi:28173257.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submission@watson.wustl.edu
----- Project Information -----
Center project name: M BB0334F11
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 201546 bases at least Q40
Consensus quality: 202453 bases at least Q30
Consensus quality: 202988 bases at least Q20
Insert size: 229000; agarose-fp
Quality coverage: 17.08 in Q20 bases; agarose-fp
Quality coverage: 11.09 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1135: contig of 1135 bp in length
* 1136 1235: gap of unknown length
* 1236 2878: contig of 1643 bp in length
* 2878 2979: gap of unknown length
* 2979 7471: contig of 4492 bp in length
* 7471 7570: gap of unknown length
* 7570 13052: contig of 5482 bp in length
* 13052 13153: gap of unknown length
* 13153 23285: contig of 10133 bp in length
* 23285 23286: gap of unknown length

```

* 23386 38432: contig of 15047 bp in length
* 38433 38532: gap of unknown length
* 38533 52633: contig of 14101 bp in length
* 52634 52733: gap of unknown length
* 52734 102557: contig of 49824 bp in length
* 102558 102657: gap of unknown length
* 102658 153484: contig of 50827 bp in length
* 153485 153584: gap of unknown length
* 153585 205202: contig of 51618 bp in length.

FEATURES

source
1..205202
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-334F11"
1..1135
/note="assembly_name:Contig8"
1236..2878
/note="assembly_name:Contig10"
2979..7470
/note="assembly_name:Contig11"
7571..13052
/note="assembly_name:Contig12"
13153..23285
/note="assembly_name:Contig13"
23386..38432
/note="assembly_name:Contig14"
38533..52633
/note="assembly_name:Contig15"
52734..102557
/note="assembly_name:Contig16"
102658..153484
/note="assembly_name:Contig17"
153585..205202
/note="assembly_name:Contig18"

ORIGIN

Query Match 77.6%; Score 19.4; DB 2; Length 205202;
Best Local Similarity 95.2%; Pred. No. 89;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AACACCTTAGTCAGATACTA 21
||| ||||| ||||| ||||| |||||
Db 78563 AACACCTTAGTCAGATACTA 78583

RESULT 9
CLPA87KB
LOCUS
Picea abies 8.7kb chloroplast DNA linear PLN 29-MAR-2001
DEFINITION
ACCESSION AJ001004
VERSION AJ001004.1 GI:2764566
KEYWORDS
acc gene; acetyl-coenzyme A carboxylase; atpB gene; atpE gene;
Hd-ATP synthase; photosystem I polypeptide I; psal gene;
ribulose-bisphosphate carboxylase large subunit; trna-Arg; trnR
gene.
SOURCE
chloroplast Picea abies (Norway spruce)
Picea abies
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.

1
Sutter, A., Philipps, A. and Wild, A.
Picea abies chloroplast genome fragment of 8,7 kb including atpE,
atpB, rbcL, trnR, accD, and psal
Unpublished
2 (bases 1 to 8694)
Sutter, A.
Direct Submission
Submitted (05-AUG-1997) Sutter A., Johannes Gutenberg-Universitaet
Mainz, Institut fuer Allgemeine Botanik, Mullerweg 6, 55099 Mainz,
GERMANY
Revised by author 24-NOV-97

FEATURES

source
1..8694
/organism="Picea abies"
/organelle="plastid:chloroplast"
/mol_type="genomic DNA"
/db_xref="taxon:3329"
/clone_lib="lambda_ZAP II"
complement (234..647)
/gene="atpB"
complement (234..647)
/gene="atpB"
/codon_start=1
/transl_table=11
/product="H+-ATP synthase subunit E"
/protein_id="CAA04457.1"
/db_xref="GI:2764567"
/db_xref="GOA:O47036"
/db_xref="SWISS-PROT:O47036"
/translation="MTLNRLVLSPNRVINDSEVKELITLSTNSGOIGVLPNHASIVAAV
DIGVMKIRLNGQSTWALMGGFAPKIDNDRITVLVNNARVDIDLKEAQETFKVAKAD
LARAEGKRAIEADVALKGRTRLEAISAPPVSN"
complement (656..2122)
/gene="atpB"
complement (656..2122)
/gene="atpB"
/codon_start=1
/transl_table=11
/product="H+-ATP synthase subunit B"
/protein_id="CAA04458.1"
/db_xref="GI:2764568"
/db_xref="GOA:O47037"
/db_xref="SWISS-PROT:O47037"
/translation="MRINLVLGVLSALVKNQVRIAGIIPVLVDVFPFPPGNMNIYNS
LIVKGGTAGQEIQVTCVQQLGNHKVRAMSDTGLTRGRMRVIDTGTAPLSPVYGG
ATLGRIFNVGEVDNLGPVDARITSPHRSAPFTLDTKLISFTGKVVLLAPY
RRGKICLFGAGVGKTVLIMELINNAKAGGVVFGVGRTRGNLDLYMDEKESG
VIDEONISKVALVYQNMPEPPGARMVGLTALTWAEYFRDVNEODVLSFIDNIERF
VQAGSEVALLGMPGSAVGYOPLATMGSLQERITSTKRGSTISCAVYVPADLTD
PAPATFAHLDTVPFSLAAGIYPADVDPDSTMLQPMVIGBEYETAGQVKQT
LQRYKELQDIILAIPLGLDEUSEEDRLIVARAKIERLFSQFFVFAEFTGPGKYVGLM
ETIRGFQMLSGLEQSFYLVGNDEATAKAMNKTES"
4172..5599
/gene="rbcL"
4172..5599
/gene="rbcL"
/codon_start=1
/transl_table=11
/product="ribulose-bisphosphate carboxylase large subunit"
/protein_id="CAA04459.1"
/db_xref="GI:2764569"
/db_xref="GOA:P48711"
/db_xref="SWISS-PROT:P48711"
/translation="MSPKTETKASVGFKAGVKDYRLTYTPEYQTKDITDILAAFRVTP
QFQVPEEAGAAVAASSTGTWTVTMDGLTSLDRYKGRCYDIEPVAGEESQFIAPVA
YPLDFEESGVNLTFTSIVGNVFGKALRALRLEDLRIPPAKTFQGPFGHGIQVERD
KINKYRPLGCTIKPLGLSAKNYGRAVVECLRGDLFTKDDENVNSOPFPMWRDRF
VFCALYKAQAEITKIKHYNLATAGTCEEMKRAVFAVELGPIVMDYLTGGFTA
NTSLAHYCRDNGLLHHRHMAVIDRQKNGHMFRLVAKALRMGGDHVGGHVTYVVGK
LEGEREITGLFVLLDRDFIEKDRSGIYFTQDMVMPGLVFAVSGGHIHWHMPALTE
IFGDDSVLQFGGGTGLHPGNAPGAVANRVALEACVQARNEGRDRLARENEVIREASK
WSPELAAACEIKWEIKFBEAVDTI"
5915..5988
/gene="trnR"
5915..5988
/gene="trnR"
/product="trna-Arg"
6244..7203
/gene="accD"
6244..7203
/gene="accD"
/codon_start=1
/transl_table=11
/product="acetyl-coenzyme A decarboxylase"

```

/protein_id="CAA04460.1"
/db_xref="GI:2764570"
/db_xref="GOA:047039"
/db_xref="SWISS-PROT:047039"
/translation="MSIRWFEDRRKIGLLKNSVERDSKDVNERNKNSIDYVKI
NRLVQCNCESLLYIRFNKSCVECYLQWNSDRIELLIDRGTRHPMEDMY
TLDVLQFCHSENPASDPLHSDESKDHTFTQIETGLTDAIQTGOLGIPALG
VNDFFMGSGMSGVGEKTRILIERATESLPVIMVACSGARMORGSPSSMOMAKIA
SALYIPQDNRLLYVSLTSPPTGGTASFGMLGDLIIABPKAYTAFAGKIVIDQTLG
QKVIDFQTEHLEFHGLFDLIVRNLLKCVLSLSELYFVLQSSS"
7765..7875
/genes="psai"
/feature="psai"
/feature="psai"
/codon_start=1
/transl_table=11
/product="Photosystem I polypeptide I"
/protein_id="CAA04461.1"
/db_xref="GI:2764571"
/db_xref="GOA:047040"
/db_xref="SWISS-PROT:047040"
/translation="MIIFNPSFFVPLVGLLPAITWVIFHLVYIQNDIF"

QY 2 ACACCTTAGTCGATCACTACTTT 25
Db 2250 ACCACCTTAGTAATCACTACTTT 2273

Query Match 76.8%; Score 19.2; DB 8; Length 8694;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

AC011752 121227 bp DNA linear PRI 30-SEP-2000
Homo sapiens BAC clone RP11-531P14 from 2, complete sequence.
AC011752 AC011752.2 GI:7534296
HTG.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Suleston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
2 (bases 1 to 121227)
Tonn,M., Drone,K. and Scott,K.
The sequence of Homo sapiens BAC clone RP11-531P14
Unpublished
3 (bases 1 to 121227)
Waterston,R.H.
Direct Submission
Submitted (13-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 121227)
Waterston,R.H.
Direct Submission
Submitted (11-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 121227)
Waterston,R.
Direct Submission
Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Apr 11, 2000 this sequence version replaced gi:6042116.
----- Genome Center

```

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0531P14

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phased quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-500G18, 200 base pair overlap. Actual start of this clone is at base position 1 of RP11-531P14; actual end is at base position 34258 of RP11-500G18.

RP11-531P14 contains a transposon in the growth of the clone, which is not part of the submitted sequence.

FEATURES

source	Location/Qualifiers
1..121227	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-531P14"
	/clone_lib="RPCI-11"
1805..2244	/rpt_family="L2"
2436..2451	/note="similar to EST AA648276 (NID:g2574705) na20d03.s1"
3307..3349	/rpt_family="L2"
4249..4270	/rpt_family="AT-rich"
5090..5111	/rpt_family="(TG)n"
5323..5538	/rpt_family="L2"
6651..6711	/rpt_family="Ricksha"
6712..6949	/rpt_family="MER2_type"
6965..7417	/rpt_family="Retroviral"
8545..8614	/rpt_family="MIR"

```

repeat_region 9587..9635
/rpt_family="L2"
repeat_region 9657..10026
/rpt_family="L2"
repeat_region 10057..10199
/rpt_family="L2"
repeat_region 11189..11353
/rpt_family="L2"
repeat_region 11388..11793
/rpt_family="L2"
repeat_region 11792..12080
/rpt_family="L2"
repeat_region 12102..12297
/rpt_family="L2"
repeat_region 12687..12790
/rpt_family="L2"
repeat_region 12790..12843
/rpt_family="L2"
repeat_region 14002..14282
/rpt_family="Retroviral"
repeat_region 14387..15337
/rpt_family="Retroviral"
repeat_region 15329..15381
/rpt_family="BUR1"
repeat_region 15403..16925
/rpt_family="MER1_type"
repeat_region 16949..17073
/rpt_family="Alu"
repeat_region 17074..17929
/rpt_family="MER1_type"
repeat_region 17930..18222
/rpt_family="Alu"
misc_feature 17930..17943
/notes="similar to EST AA554531 (NID:G2325070) nl14a02.s1"
repeat_region 18223..18354
/rpt_family="MER1_type"
repeat_region 18378..19486
/rpt_family="Retroviral"
repeat_region 19521..19992
/rpt_family="Retroviral"
repeat_region 20171..20317
/rpt_family="L2"
repeat_region 20318..20610
/rpt_family="Alu"
repeat_region 20611..20656
/rpt_family="L2"
repeat_region 23453..23586
/rpt_family="MER1_type"
repeat_region 23608..23783
/rpt_family="MIR"
repeat_region 25428..25451
/rpt_family="AT-rich"
repeat_region 25484..25599
/rpt_family="L2"
repeat_region 25974..26000
/rpt_family="L2"
repeat_region 26101..26155
/rpt_family="L2"
misc_feature 26589..26936
/notes="match to EST AI081859 (NID:G3418651) ov24c04.x1"
misc_feature 27274..27725
/notes="similar to EST AI954720 (NID:G5747030) wq32d06.x1"
repeat_region 28216..28423
/rpt_family="L2"
repeat_region 28540..28617
/rpt_family="L2"
repeat_region 29020..29096
/rpt_family="MER2_type"
repeat_region 29097..29622
/rpt_family="L1"
repeat_region 29645..30187
/rpt_family="L1"
repeat_region 30263..30288

```

```

/rpt_family="Alu"
30439..30706
/rpt_family="Alu"
30722..30879
/rpt_family="L1"
30967..31089
/rpt_family="L1"
31097..31555
/rpt_family="L1"
31555..31840
/rpt_family="L1"
33392..33438
/rpt_family="L2"
33730..33784
/rpt_family="L2"
33821..34119

Query Match 76.8%; Score 19.2; DB 9; Length 121227;
Best Local Similarity 87.5%; Pred.No.1.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 ACCACCTTAGTCAGATACCTACTTT 25
||||| ||||| ||||| |||||
Db 64479 ACCACATTAGTCAATCCTACTTT 64502

```

```

RESULT 11
AC131675 191085 bp DNA linear ROD 11-NOV-2003
LOCUS Mus musculus BAC clone RP23-328L1 from chromosome 18, complete
DEFINITION sequence.
ACCESSION AC131675
VERSION AC131675.4 GI:29925883
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Isak, A. and Bielicki, L.
1 (bases 1 to 191085)
The sequence of Mus musculus BAC clone RP23-328L1
Unpublished (2001)
REFERENCE 2 (bases 1 to 191085)
AUTHORS Wilson, R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 191085)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 191085)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (14-APR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 191085)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (14-APR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 191085)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Apr 14, 2003 this sequence version replaced gi:29244816.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
-----
Summary Statistics

```


repeat_region 16218..16285
 /rpt_family="ID"
 repeat_region 16472..16523
 /rpt_family="Alu"
 repeat_region 16697..16776
 /rpt_family="Alu"
 repeat_region 16779..16923
 /rpt_family="B4"
 repeat_region 16977..17064

Query Match 76.8%; Score 19.2; DB 10; Length 191085;
 Best Local Similarity 87.5%; Pred. No. 1.1e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACTACTT 24

Db 12117 AACAGCTTACTCAGATACTACTT 12140

RESULT 12
 AL807818/c
 LOCUS 194425 bp DNA linear VRT 18-FEB-2003
 DEFINITION Zebrafish DNA sequence from clone CH211-225H24, complete sequence.
 ACCESSION AL807818
 VERSION AL807818.14 GI:28412565
 KEYWORDS HTG.
 SOURCE Danio rerio (zebrafish)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 BEASLEY, H.
 Direct Submission
 Submitted (18-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 zfish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Feb 18, 2003 this sequence version replaced gi:28273001.

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'dr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www/Projects/D_rerio/fishmask.shtml CH211-225H24 is from a CHORI-211 BAC library VECTOR: PTARBAC2.1.

FEATURES
 source Location/Qualifiers
 1..194425
 /organism="Danio rerio"
 /mol_type="genomic DNA"

/db_xref="taxon:7955"
 /clone="CH211-225H24"
 /clone_lib="CHORI-211"

Query Match 76.8%; Score 19.2; DB 5; Length 194425;
 Best Local Similarity 87.5%; Pred. No. 1.1e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACTACTT 24

Db 165858 AACACACTTACTCAGATACTAGTT 165835

RESULT 13
 AC129307/c

LOCUS 198602 bp DNA linear HTG 25-AUG-2002
 DEFINITION Mus musculus chromosome UNK clone R24-200K10, WORKING DRAFT
 SEQUENCE, 15 unordered pieces.

ACCESSION AC129307
 VERSION AC129307.3 GI:22476079
 KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
 SOURCE Mus musculus (house mouse)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 McPherson, J.D. and Waterston, R.H.
 The sequence of Mus musculus clone
 Unpublished

REFERENCE 2 (bases 1 to 198602)
 McPherson, J.D. and Waterston, R.H.
 Direct Submission
 Submitted (28-JUL-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 3 (bases 1 to 198602)
 McPherson, J.D. and Waterston, R.H.
 Direct Submission
 Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 On Aug 25, 2002 this sequence version replaced gi:22138611.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUDSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 Contact: submissions@wustl.edu
 ----- Project Information -----
 Center project name: M BB0200K10
 ----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Chemistry: Dye-primer ET; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 193934 bases at least Q40
 Consensus quality: 195486 bases at least Q30
 Consensus quality: 196639 bases at least Q20
 Insert size: 202000; agarose-fp
 Insert size: 197629; sum-of-contigs
 Quality coverage: 12.83 in Q20 bases; agarose-fp
 Quality coverage: 11.64 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1099: contig of 1099 bp in length

```

* 1100 1199: gap of unknown length
* 1200 2248: contig of 1049 bp in length
* 2249 2348: gap of unknown length
* 2349 3711: contig of 1363 bp in length
* 3712 3811: gap of unknown length
* 3812 5920: contig of 2109 bp in length
* 5921 6020: gap of unknown length
* 6021 7194: contig of 1174 bp in length
* 7195 7294: gap of unknown length
* 7295 8673: contig of 1379 bp in length
* 8674 14281: contig of 5508 bp in length
* 14282 26301: contig of 11920 bp in length
* 26302 26401: gap of unknown length
* 26402 34164: contig of 7663 bp in length
* 34165 47001: contig of 12837 bp in length
* 47002 65387: contig of 18286 bp in length
* 65388 86344: contig of 21147 bp in length
* 86345 86734: gap of unknown length
* 86735 129492: contig of 42758 bp in length
* 129493 129592: gap of unknown length
* 129593 197861: contig of 68269 bp in length
* 197862 197962: contig of 641 bp in length.

```

FEATURES

```

Location/Qualifiers
1..198602
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-200K10"

```

```

misc_feature
1..1099
/note="assembly_name:Contig39"
misc_feature
1200..2248
/note="assembly_name:Contig45"
misc_feature
2349..3711
/note="assembly_name:Contig48"
misc_feature
3812..5920
/note="assembly_name:Contig49"
misc_feature
6021..7194
/note="assembly_name:Contig50"
misc_feature
7295..8673
/note="assembly_name:Contig51"
misc_feature
8674..14281
/note="assembly_name:Contig52"
misc_feature
14382..26301
/note="assembly_name:Contig53"
misc_feature
26402..34164
/note="assembly_name:Contig54"
misc_feature
34165..47001
/note="assembly_name:Contig55
clone_end:SP6
vector_side:left"
misc_feature
47102..65387
/note="assembly_name:Contig56"
misc_feature
65488..86634
/note="assembly_name:Contig57"
misc_feature
86735..129492
/note="assembly_name:Contig58"
misc_feature
129593..197861
/note="assembly_name:Contig59"
misc_feature
197962..198602
/note="assembly_name:Contig31"

```

ORIGIN

```

Query Match 76.8%; Score 19.2; DB 2; Length 198602;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 AACCCCTTACTGACATACACTT 24
Db 78202 AACACATTAGTCAGATACACTT 78179

```

RESULT 14

```

AC122243/c 204584 bp DNA linear ROD 04-NOV-2003
Mus musculus chromosome 17 clone RP23-148C10, complete sequence.
AC122243
AC122243.3 GI:38154054
HTG.
KEYWORDS Mus musculus (house mouse)
ORGANISM Mus musculus

```

```

REFERENCE
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
AUTHORS Wilson,R.K.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 204584)
McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Nov 4, 2003 this sequence version replaced gi:34495085.

```

FEATURES

```

Location/Qualifiers
1..204584
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="17"
/clone="RP23-148C10"

```

ORIGIN

```

Query Match 76.8%; Score 19.2; DB 10; Length 204584;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 AACCCCTTACTGACATACACTT 24
Db 143956 AACCAACAAGTCAGATACACTT 143933

```

RESULT 15

```

AC100052/c 255457 bp DNA linear HTG 12-DEC-2003
Mus musculus chromosome 8 clone RP23-32L24 map 8, *** SEQUENCING IN
AC100052
AC100052.6 GI:39752800
HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus

```

```

REFERENCE
AUTHORS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
TITLE HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
JOURNAL Mus musculus (house mouse)

```


Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 255457)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome 8, clone RP23-32L24

REFERENCE
AUTHORSJOURNAL
REFERENCE
AUTHORS

2 (bases 1 to 255457)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Katat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lechoczky,J., Levine,R., Liu,G.,
MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 255457)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepe,Y.,
Collymore,A., Cooke,A., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Katat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission

TITLE
JOURNAL
COMMENT

Submitted (12-DEC-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 12, 2003 this sequence version replaced gi:38454416.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 101651: contig of 101651 bp in length
* 101652 101751: gap of 100 bp
* 101752 112879: contig of 11128 bp in length
* 112880 112979: gap of 100 bp
* 112980 204131: contig of 91152 bp in length
* 204132 233161: gap of 100 bp
* 233162 233261: contig of 28930 bp in length
* 233262 240236: gap of 100 bp
* 240237 240397: contig of 7035 bp in length
* 240397 255457: gap of 100 bp
* 255457: contig of 15061 bp in length.

FEATURES
source

1..255457
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="8"
/map="8"
/clone="RP23-32L24"
/clone_lib="RP23-32L24" Female Mouse BAC"

ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 255457;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AACACCTTAGTCAGATACCTACTT 24

DB 93399 AAACACATTAGTCAGATACCTACTT 93376

Search completed: March 24, 2004, 23:01:26
Job time : 557.809 secs

TITLE
JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIPR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L13875
Center clone name: 32_L_24

----- Genome Center

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 18:56:32 ; Search time 130.899 Seconds
(without alignments)
811.351 Million cell updates/sec

Title: US-09-889-611A-20

Perfect score: 25

Sequence: 1 aaccaccttagtcagataactattt 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N_Geneseq_29Jan04:*
- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2003as:*
- 8: Geneseqn2003bs:*
- 9: Geneseqn2003cs:*
- 10: Geneseqn2004as:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	3	AAA71453 Human meg
2	25	100.0	128	3	AAA71434 Human meg
3	25	100.0	1431	3	AAA71435 Human meg
4	22	88.0	30	3	AAA71449 Human meg
5	20.2	80.8	6182	6	ABL34015 Human inn
6	18.8	75.2	110000	2	Continuation (17 o
7	18.6	74.4	471	4	AAK80155 Human inn
8	18.6	74.4	471	4	AAK80158 Human inn
9	18.6	74.4	471	4	AAK80159 Human inn
10	18.6	74.4	1523	3	AAC39949 Arabidops
11	18.6	74.4	2000	7	ADA73153 Rice gene
12	18.6	74.4	2999	9	ACF79501 Mouse pla
13	18.6	74.4	2999	9	ACF79502 Mouse pla
14	18.6	74.4	3053	7	ABT41815 Toxicity
15	18.6	74.4	3053	9	ACF79503 Rat plas
16	18.6	74.4	8058	6	ABN95829 Gene #232
17	18.6	74.4	24142	6	ABQ76613 C. albica
18	17.8	71.2	28564	9	ADD46508 Human gen
19	17.6	70.4	369	3	AAC28947 Human sec
20	17.6	70.4	412	6	ABL68782 Kidney ca
21	17.6	70.4	803	4	AAI94790 Human neu
22	17.6	70.4	989	4	Aah01963 Candida i
23	17.6	70.4	1074	4	AAH01569 Candida d

c	24	17.6	70.4	1637	4	AAK69081	Aak69081 Human imm
	25	17.6	70.4	1721	2	AAQ14626	Aaq14626 Human Gli
	26	17.6	70.4	1721	2	AAT34366	Aat34366 Plasmid p
	27	17.6	70.4	1721	2	AAZ32237	Aaz32237 Human gli
	28	17.6	70.4	1721	3	AAA88172	Aaa88172 pJC99 hum
	29	17.6	70.4	2435	6	ABS68447	Abs68447 DNA encod
c	30	17.6	70.4	2790	4	AAK69082	Aak69082 Human imm
	31	17.6	70.4	4708	1	AAK50415	Aan50415 Human ace
c	32	17.6	70.4	6071	9	ADB58365	Adb58365 Toxicity-
c	33	17.6	70.4	6071	9	ADB52939	Adb52939 Primary r
	34	17.4	69.6	2967	4	AAK56729	Aae56729 Human BRC
	35	17.2	68.8	5149	2	AAK32455	Aat32455 Galpain 1
	36	17.2	68.8	30967	2	AAT32454	Aat32454 Galpain 1
	37	17	68.0	480	7	ACF70412	Acf70412 Photorhab
	38	17	68.0	694	7	ACF66756	Acf66756 Photorhab
	39	17	68.0	697	7	ACF66728	Acf66728 Photorhab
c	40	17	68.0	762	7	ABX06371	S. pneumo
c	41	17	68.0	777	7	ABZ42202	Abz42202 Streptoco
c	42	17	68.0	1435	9	ADB58387	Adb58387 Toxicity-
c	43	17	68.0	1435	9	ADB52966	Adb52966 Primary r
c	44	17	68.0	2516	2	AAV52381	Aav52381 Streptoco
c	45	17	68.0	3191	4	AAF29742	Aaf29742 Castor be

ALIGNMENTS

RESULT 1
AAA71453
ID AAA71453 standard; DNA; 25 BP.
XX
AC AAA71453;
XX
DT 01-DEC-2000 (first entry)
XX
DE Human megain promoter PCR primer SEQ ID NO: 20.
XX
KW Promoter; megain; human; protein isolation; screening. PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200043528-A1.
XX
PD 27-JUL-2000.
XX
PF 25-JAN-2000; 2000WO-JP0000350.
XX
PR 25-JAN-1999; 99JP-00015667.
XX
PA (KURO/) KUROKAWA K.
PA (MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 2000-543257/49.
XX
PT DNA for promoter region of megain useful for screening proteins.
XX
PS Example 4; Page 42; 45bp; Japanese.
XX
CC This invention describes a novel DNA sequence (I) representing a promoter
CC region having part or all of a specific base sequence. The invention also
CC describes (1) a vector containing (I); (2) a cell transformed by the
CC above vector; and (3) protein produced using (I). (I) is useful for
CC screening and isolating proteins (especially transcription factors).
CC AAA71434-A71469 represent PCR primers used in the method described in the
CC invention

Sequence 25 BP; 8 A; 7 C; 2 G; 8 T; 0 U; 0 Other;

Query Match 100.0%; Score 25; DB 3; Length 25;

Best Local Similarity 100.0%; Pred. No. 0.057;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 AACACCTTACTCAGATACTACTTT 25
Db 1 AACACCTTACTCAGATACTACTTT 25

RESULT 2
AAA71434
ID AAA71434 standard; DNA; 128 BP.
XX
AC AAA71434;
XX
DT 01-DEC-2000 (first entry)
XX
DE Human megsin promoter fragment DNA.
XX
KW Promoter; megsin; human; protein isolation; screening. ss.
XX
OS Homo sapiens.
XX
FN WO200043528-A1.
XX
PD 27-JUL-2000.
XX
PF 25-JAN-2000; 2000WO-JP000350.
XX
PR 25-JAN-1999; 99JP-00015667.
XX
PA (KURO/) KUROKAWA K.
(MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 2000-543257/49.
XX
PT DNA for promoter region of megsin useful for screening proteins.
XX
PS Claim 1; Page 32; 45pp; Japanese.
XX
CC This invention describes a novel DNA sequence (I) representing a promoter
region having part or all of a specific base sequence. The invention also
describes (1) a vector containing (I); (2) a cell transformed by the
above vector; and (3) protein produced using (I). (I) is useful for
screening and isolating proteins (especially transcription factors). This
sequence represents the human megsin promoter which is described in the
method of the invention
XX
SQ Sequence 128 BP; 44 A; 29 C; 24 G; 29 T; 0 U; 2 Other;
Query Match 100.0%; Score 25; DB 3; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.07;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACACCTTACTCAGATACTACTTT 25
Db 16 AACACCTTACTCAGATACTACTTT 40

RESULT 3
AAA71435
ID AAA71435 standard; DNA; 1431 BP.
XX
AC AAA71435;
XX
DT 01-DEC-2000 (first entry)
XX
DE Human megsin promoter fragment DNA.
XX
KW Promoter; megsin; human; protein isolation; screening. ss.
XX
OS Homo sapiens.
XX
FN WO200043528-A1.

```

```

XX 27-JUL-2000.
XX
XX 25-JAN-2000; 2000WO-JP000350.
XX
XX 25-JAN-1999; 99JP-00015667.
XX
PA (KURO/) KUROKAWA K.
(MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 2000-543257/49.
XX
PT DNA for promoter region of megsin useful for screening proteins.
XX
PS Disclosure; Fig 2; 45pp; Japanese.
XX
CC This invention describes a novel DNA sequence (I) representing a promoter
region having part or all of a specific base sequence. The invention also
describes (1) a vector containing (I); (2) a cell transformed by the
above vector; and (3) protein produced using (I). (I) is useful for
screening and isolating proteins (especially transcription factors). This
sequence represents a fragment of the human megsin promoter which is
described in the method of the invention
XX
SQ Sequence 1431 BP; 466 A; 266 C; 303 G; 394 T; 0 U; 2 Other;
Query Match 100.0%; Score 25; DB 3; Length 1431;
Best Local Similarity 100.0%; Pred. No. 0.095;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACACCTTACTCAGATACTACTTT 25
Db 1319 AACACCTTACTCAGATACTACTTT 1343

RESULT 4
AAA71449
ID AAA71449 standard; DNA; 30 BP.
XX
AC AAA71449;
XX
DT 01-DEC-2000 (first entry)
XX
DE Human megsin promoter PCR primer SEQ ID NO: 16.
XX
KW Promoter; megsin; human; protein isolation; screening. PCR primer; ss.
XX
OS Homo sapiens.
XX
FN WO200043528-A1.
XX
PD 27-JUL-2000.
XX
PF 25-JAN-2000; 2000WO-JP000350.
XX
PR 25-JAN-1999; 99JP-00015667.
XX
PA (KURO/) KUROKAWA K.
(MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 2000-543257/49.
XX
PT DNA for promoter region of megsin useful for screening proteins.
XX
PS Example 5; Page 40; 45pp; Japanese.
XX
CC This invention describes a novel DNA sequence (I) representing a promoter
region having part or all of a specific base sequence. The invention also
describes (1) a vector containing (I); (2) a cell transformed by the

```

CC above vector; and (3) protein produced using (1). (1) is useful for
 CC screening and isolating proteins (especially transcription factors).
 CC AAA71434-A71469 represent PCR primers used in the method described in the
 CC invention

XX Sequence 30 BP; 12 A; 9 C; 2 G; 7 T; 0 U; 0 Other;

Query Match 88.0%; Score 22; DB 3; Length 30;
 Best Local Similarity 100.0%; Pred. No. 1.5;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACCCCTTAGTCAGATACACTAC 22
 DB 9 AACCCCTTAGTCAGATACACTAC 30

RESULT 5
 ABL34015/c
 ID ABL34015 standard; DNA; 6182 BP.

XX ABL34015;

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 1988.

KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; anianaemic; cytosatic; neutrophic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW acute myeloid leukaemia; eye disease; arteriosclerosis; anaemia;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-BP007537.

XX 30-JUN-2000; 2000DB-01032529.

XX 01-SEP-2000; 2000DE-01043826.

XX (EPG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful for
 XX diagnosis and treatment of diseases associated with abnormal cytosine
 XX methylation.

XX Claim 1; SEQ ID NO 1988; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 XX genes which are modified by the methylation of cytosines. The sequences
 XX can be used in the diagnosis and treatment of immune system disorders,
 XX including eye diseases such as retinopathy, neovascular glaucoma and
 XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 XX diseases. The present sequence is a gene of the invention

XX Sequence 6182 BP; 1934 A; 37 C; 1059 G; 3152 T; 0 U; 0 Other;

XX Query Match

XX 80.8%; Score 20.2; DB 6; Length 6182;

XX Best Local Similarity 88.0%; Pred. No. 19;

XX Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AACCCCTTAGTCAGATACACTACTTT 25
 DB 1295 AACCCCTTAGTCAGATACACTACTTT 1271

RESULT 6

AAT42063_16/c

Continuation (17 of 19) of AAT42063 from base 1600001 (Haemophilus influenzae complete g
 WP Sequence split into 19 fragments LOCUS AAT42063 Accession Aat42063

WP	Fragment Name	Begin	End
WP	AAT42063_00	1	110000
WP	AAT42063_01	100001	210000
WP	AAT42063_02	200001	310000
WP	AAT42063_03	300001	410000
WP	AAT42063_04	400001	510000
WP	AAT42063_05	500001	610000
WP	AAT42063_06	600001	710000
WP	AAT42063_07	700001	810000
WP	AAT42063_08	800001	910000
WP	AAT42063_09	900001	1010000
WP	AAT42063_10	1000001	1110000
WP	AAT42063_11	1100001	1210000
WP	AAT42063_12	1200001	1310000
WP	AAT42063_13	1300001	1410000
WP	AAT42063_14	1400001	1510000
WP	AAT42063_15	1500001	1610000
WP	AAT42063_16	1600001	1710000
WP	AAT42063_17	1700001	1810000
WP	AAT42063_18	1800001	1830121

Query Match

75.2%; Score 18.8; DB 2; Length 110000;

Best Local Similarity 90.9%; Pred. No. 1.2e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AACCCCTTAGTCAGATACACTAC 22

DB 11158 AACCCCTTAGTCAGATACACTAC 11137

RESULT 7

AAK80155

ID AAK80155 standard; DNA; 471 BP.

XX AAK80155;

XX 07-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34967.
 XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 XX cytosatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US001354.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0198974P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218230P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 18-AUG-2000; 2000US-0225759P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0227009P.
PR 01-SEP-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 05-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 25-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240360P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.

PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2000US-0254097P.
PR 05-JAN-2001; 2000US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX Disclosure; SEQ ID NO 34967; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the

CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX
XX
SQ Sequence 471 BP; 157 A; 69 C; 118 G; 127 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 4; Length 471;

Best Local Similarity 84.0%; Pred. No. 78;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AACCACTTACTGACATCTACTTT 25

Db 383 AAGACCTTGGCAGATCTACTTT 407

RESULT 8

AAK80158

ID AAK80158 standard; DNA; 471 BP.

XX AAK80158;

DT 07-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34970.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX W0200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US001354.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184664P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

XX 11-JUL-2000; 2000US-0217496P.

XX 14-JUL-2000; 2000US-0218290P.

XX 26-JUL-2000; 2000US-0220963P.

XX 14-AUG-2000; 2000US-0220964P.

XX 14-AUG-2000; 2000US-0224518P.

XX 14-AUG-2000; 2000US-0224519P.

XX 14-AUG-2000; 2000US-0225213P.

XX 14-AUG-2000; 2000US-0225214P.

XX 14-AUG-2000; 2000US-0225367P.

XX 14-AUG-2000; 2000US-0225368P.

XX 14-AUG-2000; 2000US-0225270P.

XX 14-AUG-2000; 2000US-0225447P.

XX 14-AUG-2000; 2000US-0225757P.

XX 14-AUG-2000; 2000US-0225758P.

XX 18-AUG-2000; 2000US-0225759P.

XX 22-AUG-2000; 2000US-0226799P.

XX 22-AUG-2000; 2000US-0226681P.

XX 22-AUG-2000; 2000US-0226688P.

XX 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239337P.
PR 13-OCT-2000; 2000US-0239338P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.


```

PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 74.4%; Score 18.6; DB 3; Length 1523;
Best Local Similarity 84.0%; Pred. No. 90;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACCCCTTAGTCAGATACACTCTTT 25
Db 645 AACCCCTTAGTCAGATACACTCTTT 621

RESULT 11
ADA73153
ID ADA73153 standard; DNA; 2000 BP.
AC ADA73153;
XX
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 6479.
XX
XX Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; db.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to

```

```

PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
XX gene expression.
PS Claim 27; SEQ ID NO 6479; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 475 A; 562 C; 443 G; 519 T; 0 U; 1 Other;

Query Match 74.4%; Score 18.6; DB 7; Length 2000;
Best Local Similarity 84.0%; Pred. No. 93;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACCCCTTAGTCAGATACACTCTTT 25
Db 403 AACCCCTTAGTCAGATACACTCTTT 427

RESULT 12
ACF79501
ID ACF79501 standard; cDNA; 2999 BP.
XX
XX ACF79501;
XX
XX 18-DEC-2003 (first entry)
XX
DE Mouse plasminogen activator inhibitor-1 cDNA.
XX
XX Plasminogen activator inhibitor-1; PAI-1; mouse; transgenic;
KW thrombolytic; antiasthmatic; antiinflammatory; nootropic;
KW neuroprotective; antidepressant; nephrotropic; vulnerary;
KW antiseborrheic; dermatological; antiarteriosclerotic; hepatotropic;
KW gene; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 132..1340
FT /*tag= a
FT /product= "Mouse PAI-1"
FT
XX WO2003071267-A1.
XX
XX 28-AUG-2003.
XX
PF 19-FEB-2003; 2003WO-US005008.
XX
PR 19-FEB-2002; 2002US-0358061P.
XX
PA (UYVA-) UNIV VANDERBILT.
PA (DECL/) DECLERCK P J.
XX
PI Declerck PJ, Vaughan DE, Eren M;
XX
DR WPI; 2003-721694/68.
DR P-PSDB; ABR63126.
XX
XX
PT Treating a warm-blooded vertebrate animal having a medical condition,
e.g. Alzheimer's disease, atherosclerosis, myelofibrosis or
PT glomerulosclerosis, comprises administering a plasminogen activator
inhibitor-1 inhibitor.
XX
XX Disclosure; Page 73-76; 91pp; English.
PS

```

XX The present sequence is that of cDNA encoding mouse plasminogen activator
 CC inhibitor-1 (PAI-1). The PAI-1 gene can be used to produce transgenic
 CC animals of the invention, useful for screening potential PAI-1
 CC inhibitors. Such PAI-1 inhibitors can be used to treat a warm-blooded
 CC vertebrate animal having a medical condition, e.g. alopecia, undesired
 CC weight loss, Alzheimer's disease, systemic amyloidosis, myelofibrosis,
 CC glomerulosclerosis, male pattern baldness, keloids, apocrine cysts, acne,
 CC atherosclerosis, ageing, or a wound (claimed). A method of testing a
 CC candidate composition for PAI-1 inhibition activity comprises
 CC administering the composition to a transgenic animal having a PAI-1 gene
 CC incorporated into its genome, and observing an ameliorating change in the
 CC animal indicative of inhibition of PAI-1 activity, the change being an
 CC improvement of a vascular thrombotic disorder, asthma, chronic
 CC obstructive pulmonary disease, alopecia, undesired weight loss such as
 CC anorexia, Alzheimer's disease, systemic amyloid deposition, systemic
 CC amyloidosis, myelofibrosis, glomerulosclerosis, male pattern baldness,
 CC keloids, apocrine cysts, acne, atherosclerosis, ageing, or a wound (claimed)
 CC hepatosplenomegaly, extramedullary haematopoiesis, or a wound (claimed)
 XX SQ Sequence 2999 BP; 728 A; 810 C; 743 G; 718 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 9; Length 2999;
 Best Local Similarity 84.0%; Pred. No. 98;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 AACCCCTTAGTCAGATACTACTTTT 25
 Db 2607 AACCCCTTAGTCAGATACTACTTTT 2631

RESULT 13
 ACF79502
 ID ACF79502 standard; cDNA; 2999 BP.
 XX AC ACF79502;
 XX DT 18-DEC-2003 (first entry)
 XX DE Mouse plasminogen activator inhibitor-1 cDNA.
 XX KW Plasminogen activator inhibitor-1; PAI-1; mouse; transgenic;
 XX KW thrombolytic; antiasthmatic; antiinflammatory; nootropic;
 XX KW neuroprotective; antidepressant; nephrotropic; vulnerary;
 XX KW antiseborrhoeic; dermatological; antiarteriosclerotic; hepatotropic;
 XX KW gene; ss.
 XX OS Mus musculus.
 XX PH Key Location/Qualifiers
 XX FT CDS 132..1340
 XX FT /*tag= a
 XX FT /product= "Mouse PAI-1"
 XX PN WO2003071267-A1.
 XX PD 28-AUG-2003.
 XX PF 19-FEB-2003; 2003WO-US005008.
 XX PR 19-FEB-2002; 2002US-0358061P.
 XX PA (UTVA-) UNIV VANDERBILT.
 XX PA (DECL/) DECLERCK P J.
 XX PI Declerck PJ, Vaughan DE, Eren M;
 XX WPI: 2003-721694/68.
 XX P-PSDB; ABR63127.
 XX PT Treating a warm-blooded vertebrate animal having a medical condition,
 XX e.g. Alzheimer's disease, atherosclerosis, myelofibrosis or
 XX glomerulosclerosis, comprises administering a plasminogen activator

PT inhibitor-1 inhibitor.
 XX Disclosure; Page 78-81; 91pp; English.
 XX CC The present sequence is that of cDNA encoding mouse plasminogen activator
 CC inhibitor-1 (PAI-1). The PAI-1 gene can be used to produce transgenic
 CC animals of the invention, useful for screening potential PAI-1
 CC inhibitors. Such PAI-1 inhibitors can be used to treat a warm-blooded
 CC vertebrate animal having a medical condition, e.g. alopecia, undesired
 CC weight loss, Alzheimer's disease, systemic amyloidosis, myelofibrosis,
 CC glomerulosclerosis, male pattern baldness, keloids, apocrine cysts, acne,
 CC atherosclerosis, ageing, or a wound (claimed). A method of testing a
 CC candidate composition for PAI-1 inhibition activity comprises
 CC administering the composition to a transgenic animal having a PAI-1 gene
 CC incorporated into its genome, and observing an ameliorating change in the
 CC animal indicative of inhibition of PAI-1 activity, the change being an
 CC improvement of a vascular thrombotic disorder, asthma, chronic
 CC obstructive pulmonary disease, alopecia, undesired weight loss such as
 CC anorexia, Alzheimer's disease, systemic amyloid deposition, systemic
 CC amyloidosis, myelofibrosis, glomerulosclerosis, male pattern baldness,
 CC keloids, apocrine cysts, acne, atherosclerosis, ageing, or a wound (claimed)
 CC hepatosplenomegaly, extramedullary haematopoiesis, or a wound (claimed)
 XX SQ Sequence 2999 BP; 728 A; 810 C; 743 G; 718 T; 0 U; 0 Other;

Query Match 74.4%; Score 18.6; DB 9; Length 2999;
 Best Local Similarity 84.0%; Pred. No. 98;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 AACCCCTTAGTCAGATACTACTTTT 25
 Db 2607 AACCCCTTAGTCAGATACTACTTTT 2631

RESULT 14
 ABT41815
 ID ABT41815 standard; DNA; 3053 BP.
 XX AC ABT41815;
 XX DT 26-JUN-2003 (first entry)
 XX DE Toxicity modelling related rat gene SEQ ID No 1517.
 XX KW Toxic effect; gene expression profile; renal toxicity; toxicity marker;
 XX KW database; drug screening; toxicity assay; rat; ds.
 XX OS Rattus norvegicus.
 XX PN WO200295000-A2.
 XX PD 28-NOV-2002.
 XX PF 22-MAY-2002; 2002WO-US016173.
 XX PR 22-MAY-2001; 2001US-0292335P.
 XX PR 13-JUN-2001; 2001US-0297523P.
 XX PR 19-JUN-2001; 2001US-0298925P.
 XX PR 10-JUL-2001; 2001US-0303807P.
 XX PR 10-JUL-2001; 2001US-0303808P.
 XX PR 10-JUL-2001; 2001US-0303810P.
 XX PR 28-AUG-2001; 2001US-0315047P.
 XX PR 27-SEP-2001; 2001US-0324928P.
 XX PR 22-OCT-2001; 2001US-0330462P.
 XX PR 01-NOV-2001; 2001US-0330867P.
 XX PR 21-NOV-2001; 2001US-0331805P.
 XX PR 06-DEC-2001; 2001US-0336144P.
 XX PR 19-DEC-2001; 2001US-0340873P.
 XX PR 21-FEB-2002; 2002US-0357842P.
 XX PR 21-FEB-2002; 2002US-0357843P.
 XX PR 21-FEB-2002; 2002US-0357844P.
 XX PR 15-MAR-2002; 2002US-0364134P.
 XX PR 08-APR-2002; 2002US-0370144P.

PR 08-APR-2002; 2002US-0370206P.
 PR 08-APR-2002; 2002US-0370247P.
 PR 17-APR-2002; 2002US-0372794P.
 PR 21-APR-2002; 2002US-0371679P.
 XX (GENE-) GENE LOGIC INC.
 XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
 XX WPI; 2003-148464/14.
 XX Predicting at least one toxic effect of a compound, useful for toxicity
 PT modeling, comprises preparing a gene expression profile of a tissue or
 PT cell sample exposed to the compound, and comparing the gene expression
 PT profile to a database.
 XX
 PS Example 4; Page; 446pp; English.
 XX
 CC The invention relates to a novel method of predicting at least one toxic
 CC effect of a compound. The method comprises a gene expression profile of a
 CC tissue or cell sample exposed to the compound, and comparing the gene
 CC expression profile to a database comprising at least part of the data or
 CC information given in the specification. The methods are useful for
 CC predicting at least one toxic effect of a compound, predicting the
 CC progression of a toxic effect of a compound, predicting the renal
 CC toxicity of a compound, or identifying toxicity markers in tissues or
 CC cells exposed to known renal toxin. The genes are useful as toxicity
 CC markers in drug screening and toxicity assays, in monitoring disease or
 CC physiological states, or disease progression. This polynucleotide
 CC represents a rat DNA sequence relating to the toxic effect database
 CC described in the specification. NOTE: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from the World Intellectual Property
 CC Organization
 XX
 SQ Sequence 3053 BP; 773 A; 806 C; 762 G; 712 T; 0 U; 0 Other;
 Query Match 74.4%; Score 18.6; DB 7; Length 3053;
 Best Local Similarity 84.0%; Pred. No. 98;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AACCCCTTAGTCAGATACACTCTTT 25
 DB 2665 AACCCCTTAGTCAGATACACTCTTT 2689
 RESULT 15
 ACF79503
 ID ACF79503 standard; cDNA; 3053 BP.
 XX ACF79503;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Rat plasminogen activator inhibitor-1 cDNA.
 XX
 KW Plasminogen activator inhibitor-1; PAI-1; rat; transgenic; thrombolytic;
 KW antiasthmatic; antiinflammatory; neurotropic; neuroprotective;
 KW antidepressant; nephrotropic; vulnary; antiseborrheic; dermatological;
 KW antiarteriosclerotic; hepatotropic; gene; ss.
 XX
 OS Rattus sp.
 XX
 FH Key Location/Qualifiers
 CDS 119..1327
 FT /*tag= a
 FT /product= "Rat PAI-1"
 XX
 PN WO2003071267-A1.
 XX
 PD 28-AUG-2003.
 XX
 PF 19-FEB-2003; 2003WO-US005008.

XX 19-FEB-2002; 2002US-0358061P.
 XX (UYVA-) UNIV VANDERBILT.
 PA (DECL/) DECLERCK P J.
 XX
 PI Declerck PJ, Vaughan DE, Eren M;
 XX WPI; 2003-721694/68.
 XX
 PT Treating a warm-blooded vertebrate animal having a medical condition,
 PT e.g. Alzheimer's disease, atherosclerosis, myelofibrosis or
 PT glomerulosclerosis, comprises administering a plasminogen activator
 PT inhibitor-1 inhibitor.
 XX
 PS Disclosure; Page 84-86; 91pp; English.
 XX
 CC The present sequence is that of cDNA encoding rat plasminogen activator
 CC inhibitor-1 (PAI-1). The PAI-1 gene can be used to produce transgenic
 CC animals of the invention, useful for screening potential PAI-1
 CC inhibitors. Such PAI-1 inhibitors can be used to treat a warm-blooded
 CC vertebrate animal having a medical condition, e.g. alopecia, undesired
 CC weight loss, Alzheimer's disease, systemic amyloidosis, myelofibrosis,
 CC glomerulosclerosis, male pattern baldness, keloids, apocrine cysts, acne,
 CC atherosclerosis, ageing, or a wound (claimed). A method of testing a
 CC candidate composition for PAI-1 inhibition activity comprises
 CC incorporating the composition to a transgenic animal having a PAI-1 gene
 CC incorporated into its genome, and observing an ameliorating change in the
 CC animal indicative of inhibition of PAI-1 activity, the change being an
 CC improvement of a vascular thrombotic disorder, asthma, chronic
 CC obstructive pulmonary disease, alopecia, undesired weight loss such as
 CC anorexia, Alzheimer's disease, systemic amyloid deposition, systemic
 CC amyloidosis, myelofibrosis, glomerulosclerosis, male pattern baldness,
 CC keloids, apocrine cysts, acne, atherosclerosis, ageing,
 CC hepatosplenomegaly, extramedullary haematopoiesis, or a wound (claimed)
 XX
 SQ Sequence 3053 BP; 773 A; 806 C; 762 G; 712 T; 0 U; 0 Other;
 Query Match 74.4%; Score 18.6; DB 9; Length 3053;
 Best Local Similarity 84.0%; Pred. No. 98;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AACCCCTTAGTCAGATACACTCTTT 25
 DB 2665 AACCCCTTAGTCAGATACACTCTTT 2689
 Search completed: March 24, 2004, 22:28:23
 Job time : 134.899 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 21:57:32 ; Search time 1217.13 Seconds
(without alignments)
613.370 Million cell updates/sec

Title: US-09-889-611A-20

Perfect score: 25
Sequence: 1 accacccttagtcagataactattt 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estba.*

2: em_esthum.*

3: em_estlin.*

4: em_estmus.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_estc.*

9: gb_est1.*

10: gb_est2.*

11: gb_est3.*

12: gb_est4.*

13: gb_est5.*

14: gb_est6.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_pbg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.2	80.8	870	13	BX783787
2	19.8	79.2	467	28	AQ586782
3	19.4	77.6	549	9	AI188862
4	19.2	76.8	244	9	AI349205

5	18.8	75.2	569	14	CF793467
6	18.6	74.4	319	14	H85044
7	18.6	74.4	364	13	BY171910
8	18.6	74.4	377	14	H85041
9	18.6	74.4	391	9	AA849009
10	18.6	74.4	406	13	BY514030
11	18.6	74.4	411	13	BY688349
12	18.6	74.4	413	13	BY573260
13	18.6	74.4	414	13	BY513444
14	18.6	74.4	417	13	BY503440
15	18.6	74.4	418	12	BM207812
16	18.6	74.4	420	13	BY488597
17	18.6	74.4	431	10	BB823661
18	18.6	74.4	432	10	BB820649
19	18.6	74.4	438	10	BB819721
20	18.6	74.4	439	13	BY504295
21	18.6	74.4	441	10	AW916277
22	18.6	74.4	441	10	BF282272
23	18.6	74.4	449	13	BY393529
24	18.6	74.4	450	10	BB828056
25	18.6	74.4	454	13	BY591869
26	18.6	74.4	460	14	CB737667
27	18.6	74.4	470	9	AA251520
28	18.6	74.4	503	9	AA915138
29	18.6	74.4	512	9	AI180309
30	18.6	74.4	532	13	BU760030
31	18.6	74.4	556	10	BF556719
32	18.6	74.4	565	12	BG922285
33	18.6	74.4	661	14	CB233697
34	18.6	74.4	722	14	CB571286
35	18.6	74.4	724	28	BH145327
36	18.6	74.4	742	28	BZ107623
37	18.6	74.4	874	29	CNS0712R
38	18.6	74.4	883	10	BE379402
39	18.6	74.4	924	28	AZ139264
40	18.6	74.4	1126	28	CC204482
41	18.4	73.6	485	14	CB220841
42	18.4	73.6	655	14	CB444113
43	18.4	73.6	3764	11	AK087514
44	18.2	72.8	236	10	BF552977
45	18.2	72.8	273	10	AW992361

ALIGNMENTS

RESULT 1
BX783787

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

UNPUBLISHED (2003)

COMMENT

870 bp mRNA linear EST 10-DEC-2003
BX783787 XGC-egg Silurana tropicalis cDNA clone Tegg059102 3', mRNA
sequence.
BX783787
BX783787.1 GI:39690993
EST.
Silurana tropicalis (western clawed frog)
Silurana tropicalis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Silurana.
1 (bases 1 to 870)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: Tegg059102.q1kt7
Sequencing primer: T7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from egg.

CF793467 888857 MA
H85044 YS66B06.s1
BY171910 BY171910
H85041 YS66A07.s1
AA849009 EST191771
BY514030 BY514030
BY688349 BY688349
BY573260 BY573260
BY513444 BY513444
BY503440 BY503440
BM207812 C0618D02-
BY488597 BY488597
BB823661 BB823661
BB820649 BB820649
BB819721 BB819721
BY504295 BY504295
AW916277 EST147581
BF282272 EST146977
BY393529 BY393529
BB828056 BB828056
BY591869 BY591869
CB737667 AMGNNUC.M
AA251520 Z504F02.F
AA915138 Z503H12.F
AI180309 EST224052
BU760030 UI-R-FS1-
BF556719 UI-R-CO-g
BG922285 602820076
CB233697 UI-R-DVO-
CB571286 AGENCOURT
BH145327 TDGEV23PTH
BZ107623 CH230-134
AL424393 T3 end of
BE379402 601237183
AZ139264 SP 0177 B
CC204482 CH261-26E
CB220841 LAB027A11
CB444113 695241 MA
AK087514 Mus muscu
BF552977 UI-R-C2-n
AW992361 CM3-BN002


```

VERSION      AI349205.1  GI:4086411
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 244)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL      Tumor Gene Index
COMMENT      Unpublished (1997)
             Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Herbert Morse, M.D., Michael R. Emmert-Buck,
             M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html
             Insert Length: 347 Std Error: 0.00
             Seq primer: -40UP from Gibco.
FEATURES     Location/Qualifiers
             1..244
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="IMAGE:2049773"
             /tissue_type="stem cell 34+/38+"
             /dev_stage="adult"
             /lab_host="DH10B"
             /clone_lib="NCI CGAP HSC2"
             /note="Organ: bone marrow; Vector: pAMP1; mRNA made from
             bone marrow; stem cells 34+/38+, cDNA made by oligo-dT
             priming. Directionally cloned. Size-selected on agarose
             gel, average insert size 400 bp. Primary library,
             non-amplified."
ORIGIN
Query Match 76.8%; Score 19.2; DB 9; Length 244;
Best Local Similarity 87.5%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACCACCTTAGTCAGATCTACTTT 25
||||| ||||||| |||||
Db 104 ACCACCTTGTGCAGATCTGTTT 127

RESULT 5
CF793467 569 bp mRNA linear EST 21-OCT-2003
LOCUS
DEFINITION 888857 MARC 4PTG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION CF793467
VERSION
KEYWORDS
SOURCE
ORGANISM Sus scrofa (pig)
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 569)
Smith, T.P.L., Freking, B.A., Ford, J.J., Vallet, J.L., Wise, T.A.,
Noneman, D.J., Wray, J.E. and Keele, J.W.
Porcine EST collection using a normalized library constructed from
embryos representing early developmental stages
Unpublished (2003)
CONTACT: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and

```

```

trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: TMW8011 row: M column: 8
Seq primer: GTAATACGACTACTACTAGG.
FEATURES     Location/Qualifiers
             1..569
             /organism="Sus scrofa"
             /mol_type="mRNA"
             /db_xref="taxon:9823"
             /tissue_type="pooled"
             /lab_host="DH10B"
             /clone_lib="MARC 4PTG"
             /note="Vector: pCDNA3.1; Site_1: EcoRI; Site_2: NotI;
             Library made with combined RNA from day-10, day-13,
             day-15, day-25, and day-30 whole embryos."
ORIGIN
Query Match 75.2%; Score 18.8; DB 14; Length 569;
Best Local Similarity 90.9%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATCTACT 22
||||| ||||||| |||||
Db 33 AACCACTTAGTCAGATCTACT 54

RESULT 6
H85044/c
LOCUS
DEFINITION H85044.1 GI:1063699
ACCESSION H85044
VERSION
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 319)
Hillier, L., Clark, N., Dubucq, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohldmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 299
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: Promega -21m13
High quality sequence stop: 299.
FEATURES     Location/Qualifiers
             1..319
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="GDB:384772"
             /db_xref="taxon:9606"
             /clone="IMAGE:219731"
             /sex="male"
             /tissue_type="retina"
             /dev_stage="55 year old"
             /lab_host="DH10B (ampicillin resistant)"
             /clone_lib="Soares retina N2b4HR"
             /note="Organ: eye; Vector: pT7n3D (Pharmacia) with a
             modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

```

strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p7T3 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)⁺ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Patima Bonaldo. "

ORIGIN

Query Match 74.4%; Score 18.6; DB 14; Length 319;
Best Local Similarity 84.0%; Pred. No. 6.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACACTTTT 25

Db 139 AGCCACCTTAATCAGATACACTTTT 115

RESULT 7
BY171910 364 bp mRNA linear EST 10-DEC-2002
LOCUS
DEFINITION
BY171910 RIKEN full-length enriched, bone marrow macrophage Mus
musculus cDNA clone I830090J19 5', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Mus musculus (house mouse)
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE
AUTHORS

1 (bases 1 to 364)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Shonobach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.I., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wyshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayase, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, S., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

Fax: 81-45-503-9216
Email: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Omo, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by David A. Hume (Depts. of Biochemistry
and Microbiology/Parasitology Institute for Molecular Bioscience
University of Queensland Brisbane, Q 4072 Australia) whose
assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

FEATURES
source

Location/Qualifiers
1..364
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="I830090J19"
/tissue_type="bone marrow"
/cell_type="macrophage"
/clone_lib="RIKEN full-length enriched, bone marrow
macrophage"

ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 364;
Best Local Similarity 84.0%; Pred. No. 6.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACACTTTT 25

Db 28 AACCACTTAGTCAGATACACTTTT 52

RESULT 8

BY171910/c

LOCUS

DEFINITION
H85041 y566a07.s1 Soares retina N2B4HR Homo sapiens cDNA clone
IMAGE:219732 3' similar to contains Alu repetitive element; contains
PTR5 repetitive element ; mRNA sequence.

ACCESSION

H85041

KEYWORDS

EST.

SOURCE

ORGANISM

Homo sapiens (human)

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 377)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

TITLE

The WashU-Merck EST Project

JOURNAL
COMMENT

Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

High quality sequence stops: 370
 Email: estewatson.wustl.edu
 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 590 Std Error: 0.00
 Seq primer: Promega -21ml3
 High quality sequence stops: 370.

FEATURES
SOURCE

```

1. 377
/organisms="Homo sapiens"
/mol_type="mrna"
/db_xref="GDB:3847725"
/db_xref="taxon:9606"
/clone="IMAGE:219732"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (Ami)

```

/lab_host="PH105 (ampicillin resistant)"
 /clone_lib="Soares retina N2b4HR"
 /notes="Organ: eye; Vector: pT773D (Pharmacia) with a
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5',
 TGTTACCATCTGAGTGGAGCGCCGCTTTTTTTTTTTTTTTT 3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT773 vector
 (Pharmacia). The retinas were obtained from a 55 year old
 Caucasian and total cellular poly(A)⁺ RNA was extracted 6
 hrs after their removal. The retina RNA was kindly
 provided by Roderick R. McInnes M.D. Ph.D. from the
 University of Toronto. Library constructed by Bento
 Soares and M. Fatima Ronaldo."

ORIGIN

Query Match 74.4%; Score 18.6; DB 14; Length 377;
Best Local Similarity 84.0%; Pred. No. 6.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0

Qy 1 AACCACTTAGTCAGATACTTT 25
Db 178 AGCCACCTAACTCAGATACTTT 15

RESULT 9

AA849009/c	AA849009	391 bp	linear	EST 30-APR-1998
LOCUS	EST191771	Normalized rat lung, Bento Soares	Rattus sp.	cDNA clone
DEFINITION	RIJUA196 3' end, mRNA sequence.			

ACCESSION AA849009

VERSION AA849009.1 GI:293

KEYWORDS EST.

SOURCE	Rattus sp.
ORGANISM	Rattus sp.

ORGANISM Rattus sp.
Eukaryota: Metazoa

Eukaryota, Metazoa:

Rattus.

REFERENCE 1 (bases 1 to 391)

AUTHORS Lee, N.H., Glodek,

Kerlavage, A.R. and

Rat Genome Project

Gene Index

BOOKNAME
SURPRISED (1998)
CONTACT: Lee NH
COMMENT

CONTACT: DEE, MI
The Institute for

9712, Medical Cent

Tel: (301)-838-3522

FEATURES source

Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1..391
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="ATCC (inhost):208037"
/db_xref="taxon:10118"
/clone="RLUA196"
/clone_lib="Normalized rat lung, Bento Soares"
/note="Organ: lung; Vector: pT7T3Pac; Site_1: EcoRI;
Site_2: NotI"

ORIGIN

Query Match 74.4%; Score 18.6; DB 9; Length 391;
Best Local Similarity 84.0%; Pred. No. 6.7e+02;
Matches 21: Conservative 0: Mismatches 4: Indels

QY 1 AACCACTTAGTCAGATACTTT 25
|||
364 AACCACTTAGTTAGATAATCTTT 34
|||

RESULT 10

BY514030	BY514030	406 bp	mRNA	linear	EST 14-DEC-2002
LOCUS	BY514030	RIKEN full-length enriched, bone marrow macrophage Mus musculus cDNA clone I830153E12 3', mRNA sequence.			
DEFINITION	BY514030				
ACCESSION	BY514030				
VERSION	BY514030.1	GI:26848409			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
	1 (bases 1 to 406)				
REFERENCE	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,				
AUTHORS					

Schönbach, C., Gojbori, T., Baldarelli, R., Hill, D. P., Bult, C.,
Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,
Batalov, S., Baisel, K. W., Blake, J. A., Bradt, D., Brusic, V.,
Chothia, C., Corbani, L. E., Cousins, S., Dalla, E. S., Dragani, T. A.,
Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustinich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedierski, R. M., King, B. L., Konagaya, A.,
Kurochkin, I. V., Lee, Y., Lennard, B., Lyons, P. A., Maglott, D. R.,
Malkais, I., Marchionni, L., McKenzie, L. L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J. C., Reed D. J., Reid, J., Ring, B. Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempile, C. A., Setou, M. K., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, K.,
Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., waki, K., Kawai, J., Mizawa, K.,
Arawaka, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, K., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Landier, E. S.,
Rogers, J., Birney, E. and Havaashizaki, Y.

TITLE

JOURNAL	of 60,770 full-length cDNAs
MEDLINE	Nature 420, 563-573 (2002)
PUBMED	22354683
COMMENT	12456851

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@genome.gsc.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Inotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by David A. Hume (Depts. of Biochemistry
 and Microbiology/Parasitology Institute for Molecular Bioscience
 University of Queensland Brisbane, Q 4072 Australia) whose
 assistance we gratefully acknowledge.
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

FEATURES

source

Location/Qualifiers

```

1. 406
  /organism="Mus musculus"
  /mol_type="mRNA"
  /strain="C57BL/6J"
  /db_xref="taxon:10090"
  /clone="I830153E12"
  /tissue_type="bone marrow"
  /cell_type="macrophage"
  /clone_lib="RIKEN full-length enriched, bone marrow
  macrophage"

```

ORIGIN

```

Query Match      74.4%; Score 18.6; DB 13; Length 406;
Best Local Similarity 84.0%; Pred. No. 6.8e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 AACCACTTAGTCAGATACACTTTT 25
    |||||
Db 12 AACCACTTAGTCAGATACACTTTT 36

```

RESULT 11

BY688349

LOCUS

```

BY688349 BY688349 411 bp mRNA linear EST 16-DEC-2002
clone I320028N18 3', mRNA sequence.

```

ACCESSION

BY688349

VERSION

BY688349.1 GI:27093998

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 411)

```

REFERENCE

AUTHORS

```

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yananaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,

```

Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bruscia, V.,
 Chochia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, I.A.,
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
 Guinrich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
 Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
 Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,
 Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
 Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Inotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
 Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-res@genome.gsc.riken.go.jp/

URL: http://genome.gsc.riken.go.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Inotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
 Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Takashi Ishikawa (Department of Surgery
 2 Yokohama City University 3-9 Fukuura, Kanazawa-ku, Yokohama
 236-0004 Japan) whose assistance we gratefully acknowledge.
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

FEATURES

source

Location/Qualifiers

1. 411

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="I320028N18"

/cell_type="stroma cell"

/clone_lib="RIKEN full-length enriched, stroma cell"

ORIGIN

```

Query Match      74.4%; Score 18.6; DB 13; Length 411;
Best Local Similarity 84.0%; Pred. No. 6.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACACTCTTT 25
    |||||
Db 16 AACCACTTAGTCAGATACACTCTTT 40

RESULT 12
LOCUS BY573260
DEFINITION BY573260 RIKEN full-length enriched, activated spleen Mus musculus
ACCESSION BY573260
VERSION BY573260.1 GI:26908442
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
1 (bases 1 to 413)
AUTHORS Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Oatono, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Kurochi, H., Kwasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kawaji, H., Kawasawa, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Perlea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavoian, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,
Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
DOI: 10.1038/420563a
PubMed 12466851

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to

```

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers
1..413
/organism="Mus musculus"
/mol_type="mRNA"
/strain="M0D"
/db_xref="taxon:10090"
/clone="F830301I21"
/tissue_type="activated spleen"
/clone_lib="RIKEN full-length enriched, activated spleen"

ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 413;
Best Local Similarity 84.0%; Pred. No. 6.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACCACTTAGTCAGATACACTCTTT 25
|||||
Db 27 AACCACTTAGTCAGATACACTCTTT 51
|||||

RESULT 13

BY513444
LOCUS BY513444
DEFINITION BY513444 RIKEN full-length enriched, bone marrow macrophage Mus musculus cDNA clone 1830149K22 3', mRNA sequence.

ACCESSION

BY513444
VERSION BY513444.1 GI:26847823

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 414)

AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Oatono, N., Saito, R., Suzuki, H., Yamanaka, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Kurochi, H., Kwasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
Kawaji, H., Kawasawa, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W.J., Perlea, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J.C., Reid, J., Ring, B.Z., Ringwald, M.,
Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavoian, M., Zhu, Y., Zimmer, A., Carninci, P.,

Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by David A. Hume (Depts. of Biochemistry and Microbiology/Parasitology Institute for Molecular Bioscience University of Queensland Brisbane, Q 4072 Australia) whose assistance we gratefully acknowledge.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers

1..414

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clones="I830148K22"

/tissue_type="bone marrow"

/cell_type="macrophage"

/clone_lib="RIKEN full-length enriched, bone marrow macrophage"

ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 414;
Best Local Similarity 84.0%; Pred. No. 6.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AACCACTTACTGACTACTTCTT 25
|||||
Db 20 AACCACTTACTGACTACTTCTT 44
|||||

RESULT 14
BY503440
LOCUS

417 bp mRNA linear EST 14-DEC-2002

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

BY503440 RIKEN full-length enriched, bone marrow macrophage Mus musculus cDNA clone I830081H09 3', mRNA sequence.
BY503440
BY503440.1 GI:26837819
EST.
Mus musculus (house mouse)
Mus musculus
Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 417)
Okazaki, Y., Furuno, M., Kaakawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yananaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Shosbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Ciothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grilmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierzki, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Resole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L.G., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyshew-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
PUBMED

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by David A. Hume (Depts. of Biochemistry and Microbiology/Parasitology Institute for Molecular Bioscience University of Queensland Brisbane, Q 4072 Australia) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

```
1. .417
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="T830081H09"
/tissue_type="bone marrow"
/cell_type="macrophage"
/clone_lib="RIKEN full-length cDNA library"
/clone_phage="macrophage"
```

ORIGIN

```

Query Match      74.4%; Score 18.6; DB 13; Length 417;
Best Local Similarity 84.0%; Pred. No. 6.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

1 AACCACTTAGTCAGATACTTT 25
|||
20 AACCACTTAGTCAGATACTTT 44

RESULT 15	BM207812	418 bp	mRNA	linear	EST 08-JUN-2003
NEW207812/c	C0618D02-3	NIA Mouse Trophoblast Stem Cell	cDNA Library	(Long)	Mus
LOCUS	Musculus	cDNA clone NIA:C0618D02	IMAGE:30021925	3', mRNA	sequence.
DEFINITION	BM207812				
ACCESSION	BM207812.2	GI:31531823			
VERSION					
KEYWORDS					
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1. (bases 1 to 418)				
AUTHORS	Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.				
TITLE	Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification				

FEATURES
SOURCE

```

Location/Qualifiers
1. .418
  /organism="Mus musculus"
  /mol_type="mRNA"
  /strain="B6J.EGFP transgenic ICR mice"
  /db_xref="NIA:GSC:C0618D02-3"
  /db_xref="taxon:10090"
  /clone="NIA.C0618D02 IMAGE:30021925"
  /tissue_type="trophoblast stem cell"
  /dev_stage="3.5-dpc"
  /lab_host="PH108"
  /clone_lib="NIA Mouse Trophoblast Stem
  (long)"
  /vector="psPORT1 (Invitrogen): 5

```

NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (<http://lgaun.grc.nia.nih.gov/cDNA>). This is a long-transcript enriched cDNA library (Ref. Genomes Res. 11: 1553-1558 (2001). [PMID: 11544193]). Total RNAs were obtained from Dr. Janet Rossant and Tilo Kunath (Samuel Lunenfeld Research Institute, Canada). Double-stranded cDNAs were synthesized with an Oligo(dT) primer [5'-pGACGTAGTCTAGATCGGCGGCCCTTTTTTTT-3'] from 4 µg of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORI plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.6 kb. The library was constructed by Yulan Piao (NIA).

ORIGIN

```
Query Match      74.4%; Score 18.6; DB 12; Length 418;
Best Local Similarity 84.0%; Pred. No. 6.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

1 AACCACTTAGTCAGATACTTT 25
394 AACCACTTAGTAGATAATCTTT 370

Search completed: March 25, 2004, 00:13:50
Job time : 1222.13 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 21:53:36 ; Search time 906.607 Seconds
(without alignments)
1960.127 Million cell updates/sec

Title: US-09-889-611A-60_COPY_2_42

Perfect score: 41

Sequence: 1 gaagaaactaatacaacc.....ccttagtcagataactacttt 41

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sv.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.hgo.hum.*

40: em.hgo.mus.*

41: em.hgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	39.4	96.1	4229	9	AF234618	AF234618 Homo sapi
2	39.4	96.1	157284	9	AC072051	AC072051 Homo sapi
3	39.4	96.1	189092	2	AC027524	AC027524 Homo sapi
4	33.6	82.0	6182	6	AX346917	AX346917 Sequence
5	29.4	71.7	207510	2	AC147488	AC147488 Otomur
6	28.2	68.8	198669	10	AC103453	AC103453 Rattus no
7	28.2	68.8	219071	2	AC133259	AC133259 Rattus no
8	25	61.0	6182	6	AX346916	AX346916 Sequence
9	25	61.0	121538	10	AC129295	AC129295 Mus muscu
10	25	61.0	200844	2	AC134414	AC134414 Mus muscu
11	25	61.0	213457	2	AC110569	AC110569 Mus muscu
12	24.6	60.0	160615	10	AL672023	AL672023 Mouse DNA
13	24	58.5	3473	9	HSN804835	AL833522 Homo sapi
14	24	58.5	5276	6	AX251133	AX251133 Sequence
15	24	58.5	5276	6	AX345728	AX345728 Sequence
16	24	58.5	146746	9	AL353718	AL353718 Human DNA
17	24	58.5	222097	2	AC129874	AC129874 Rattus no
18	24	58.5	228783	2	AC094377	AC094377 Rattus no
19	24	58.5	235965	2	AC127103	AC127103 Rattus no
20	23.6	57.6	148648	9	AC021713	AC021713 Homo sapi
21	23.4	57.1	99539	8	AC113332	AC113332 Oryza sat
22	23.4	57.1	122599	8	AC130605	AC130605 Oryza sat
23	23.4	57.1	161161	5	AL935146	AL935146 Zebrafish
24	23.4	57.1	165700	10	AC132342	AC132342 Mus muscu
25	23.4	57.1	233830	2	AC141892	AC141892 Mus muscu
26	23.4	57.1	289080	2	AC111582	AC111582 Rattus no
27	23.2	56.6	1005	3	AY005308	AY005308 Xylocopa
28	23.2	56.6	132093	2	AC141964	AC141964 Rattus no
29	23.2	56.6	260121	2	AC130738	AC130738 Rattus no
30	23.2	56.6	338614	2	AC103268	AC103268 Rattus no
31	23	56.1	4038	9	HS24P521	X67640 H.sapiens H
32	23	56.1	69037	10	AC003949	AC003949 Mus muscu
33	23	56.1	116103	9	AL590653	AL590653 Human DNA
34	23	56.1	123851	9	AC113410	AC113410 Homo sapi
35	23	56.1	131886	2	AC020834	AC020834 Mus muscu
36	23	56.1	158296	2	BX530098	BX530098 Danio rer
37	23	56.1	198694	10	BX072557	BX072557 Mouse DNA
38	23	56.1	216921	2	BX322658	BX322658 Mus muscu
39	23	56.1	235952	2	BX784023	BX784023 Danio rer
40	23	56.1	346601	2	AC111981	AC111981 Rattus no
41	22.8	55.6	168169	2	AC145953	AC145953 Pan trogl
42	22.8	55.6	196870	2	AC115209	AC115209 Rattus no
43	22.8	55.6	201286	2	AC102228	AC102228 Mus muscu
44	22.8	55.6	219817	2	AC099196	AC099196 Rattus no
45	22.8	55.6	220129	2	AC096823	AC096823 Rattus no

ALIGNMENTS

RESULT 1
AF234618
LOCUS AF234618 4229 bp DNA linear PRI 29-MAR-2002
DEFINITION Homo sapiens megain gene, promoter region and partial sequence.
ACCESSION AF234618
VERSION AF234618.2 GI:19808130
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4229)
AUTHORS Miyata,T., Nangaku,M., Inagi,R. and Kurokawa,K.
TITLE Transcriptional regulation of a mesangium-predominant gene, megain
JOURNAL Unpublished

```

REFERENCE
AUTHORS
TITLE
JOURNAL
1 (bases 1 to 4229)
Miyata, T., Nangaku, M., Inagi, R. and Kurokawa, K.
Direct Submission
Submitted (15-FEB-2000) Institute of Medical Sciences and
Department of Internal Medicine, Tokai University School of
Medicine, Bohseidai, Isehara, Kanagawa 259-1193, Japan
2 (bases 1 to 4229)
Miyata, T., Nangaku, M., Inagi, R. and Kurokawa, K.
Direct Submission
Submitted (29-MAR-2002) Institute of Medical Sciences and
Department of Internal Medicine, Tokai University School of
Medicine, Bohseidai, Isehara, Kanagawa 259-1193, Japan
REMARK
COMMENT
Sequence update by submitter
On Mar 29, 2002 this sequence version replaced gi:18000453.
FEATURES
Location/Qualifiers
1..4229
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21.3"
1..4021
promoter
4022..4229
mRNA
/product="megasin"
ORIGIN
Query Match 96.1%; Score 39.4; DB 9; Length 4229;
Best Local Similarity 97.6%; Pred. No. 7.4e-05;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GAATGAACATCAATCAACACCACTTACTGACACTACTTT 41
Db 3920 GAATGAACATCAATCAACACCACTTACTGACACTACTTT 3960
RESULT 2
AC072051/c 157284 bp DNA linear PRI 03-DEC-2001
LOCUS
AC072051 Homo sapiens chromosome, clone RP11-79D21, complete sequence.
DEFINITION
AC072051 Homo sapiens
ACCESSION
AC072051.8 GI:16974280
VERSION
AC072051.8
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 157284)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome, clone RP11-79D21
JOURNAL
Unpublished
AUTHORS
REFERENCE
1 (bases 1 to 157284)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
Anderson, S., Balding, J., Barna, N., Bastien, V., Bedalov, F.,
Boguslavskiy, L., Bouckhgalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRocque, K., Lamazares, R., Lander, E., Levine, R., Liu, G.,
Levine, R., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, J., Peterson, K., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Wu, X., Wyman, D., Ye, W.J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
FEATURES
Location/Qualifiers
1..157284
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-79D21"
TITLE

```

JOURNAL REFERENCE AUTHORS

Submitted (07-JUN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 157284)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Bouckhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Fargo, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Lander, E., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL REFERENCE AUTHORS

Direct Submission

Submitted (25-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 157284)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Bouckhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Fargo, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Lander, E., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL COMMENT

Submitted (03-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.P.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RW/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research Center code: WIBR Web site: http://www-seg.wi.mit.edu Contact: sequence_submissions@genome.wi.mit.edu ----- Project Information Center project name: L966 Center clone name: 79_D_21 ----- Location/Qualifiers 1..157284 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /clone="RP11-79D21"

FEATURES source

Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Gord, S., Goyette, M., Graham, D., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menes, L., Minova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 1, 2002 this sequence version replaced gi:11990731.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WlBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8516
Center clone name: 317_G1

* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 85659: contig of 85659 bp in length
* 85660: gap of 100 bp
* 85760 163014: contig of 77255 bp in length
* 163015 163114: gap of 100 bp
* 163115 189092: contig of 25978 bp in length.

Location/Qualifiers
1..189092
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/map="18"
/clone="RP11-317G1"
/clone_lib="RP11-317G1"
/clone="RP11-317G1"
/clone="RP11-317G1"

Query Match 96.1%; Score 39.4; DB 2; Length 189092;
Best Local Similarity 97.6%; Pred. No. 5.4e-05;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAATGAACCTACATACCAACCCCTTAGTCAGATACCTATTT 41
|||||
DB 66563 GAATGAACCTACATACCAACCCCTTAGTCAGATACCTATTT 66603

RESULT 4
AX346917/c
LOCUS AX346917
DEFINITION Sequence 1988 from Patent WO0200928.
ACCESSION AX346917

VERSION AX346917.1 GI:18494803
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1988 03-JAN-2002;
EpiGenomics AG (DE)
FEATURES
source
1..6182
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN
Query Match 82.0%; Score 33.6; DB 6; Length 6182;
Best Local Similarity 90.0%; Pred. No. 0.014; Indels 0; Gaps 0;
Matches 36; Conservative 0; Mismatches 4;

CY 2 AATGAACCTACATACCAACCCCTTAGTCAGATACCTATTT 41
|||||
DB 1310 AATGAACCTACATACCAACCCCTTAGTCAGATACCTATTT 1271

RESULT 5
AC147488/c
LOCUS AC147488
DEFINITION Ocolemur garnettii clone CH256-273K16, WORKING DRAFT SEQUENCE, 2
unordered pieces.
ACCESSION AC147488
VERSION AC147488.1 GI:38638706
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Ocolemur garnettii (small-eared galago)
ORGANISM Ocolemur garnettii

REFERENCE 1 (bases 1 to 207510)
AUTHORS Antonellis, A., Ayele, K., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Chu, G., Coleman, B., Coleman, H., Engle, J., Granite, S., Guan, X., Gupta, J., Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Hu, P., Hurlb, B., Idol, J.R., Karlins, E., Kwong, P., Latic, P., Larson, S., Lee-Edin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B., Margulies, E.H., Masello, C., Maskeri, B., McDowell, J., Mullikin, J.C., Paguirigan, C., Pearson, R., Portnoy, M.E., Prasad, A., Reddix-Dugue, N., Schandier, K., Schueler, M.G., Shah, K., Sison, C., Stantropop, S., Thomas, J.W., Thomas, P.J., Tsipouri, V., Vogt, J.L., Wetherby, K.D., Young, A. and Green, E.D.

NISC Comparative Sequencing Initiative

Unpublished

2 (bases 1 to 207510)

Green, E.D.

Direct Submission

Submitted (03-DEC-2003) NIH Intramural Sequencing Center, 8717

Grovenmont Circle, Gaithersburg, MD 20877, USA

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc_zoo@nhgri.nih.gov

----- Project Information

Center project name: fhy

Center clone name: 273K16

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 207213 bases at least Q40

Consensus quality: 207263 bases at least Q30

Consensus quality: 207289 bases at least Q20

Insert size: 218000; agarose-gel
 Insert size: 207410; sum-of-contigs
 Quality coverage: 14.36x in Q20 bases; agarose-gel
 Quality coverage: 15.10x in Q20 bases; sum-of-contigs

***** NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 38181: contig of 38181 bp in length
 * 38182 38281: gap of unknown length
 * 38282 207510: contig of 169229 bp in length.

FEATURES

source

Location/Qualifiers
 1. 207510
 /organism="Otolemur garnettii"
 /mol_type="genomic DNA"
 /db_xref="taxon:30611"
 /clone="CH256-273K16"
 /clone_lib="CH256"

misc_feature

1. 38181
 /notes="assembly_fragment"

clone_end:T7
 vector_side:left

misc_feature

38282..207510
 /notes="assembly_fragment"

clone_end:SP6
 vector_side:right

ORIGIN

Query Match 71.7%; Score 29.4; DB 2; Length 207510;
 Best Local Similarity 84.6%; Pred. No. 0.44;
 Matches 33; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GAATGAACATACATACACACCTTGTAGTCAGACTACT 39
 |||||
 Db 151706 GAATGAACCCATATGATGCTTGTAGTCAGACTACT 151668

RESULT 6

AC103453

LOCUS

AC103453 199669 bp DNA linear ROD 07-JUN-2003
 Rattus norvegicus 13 BAC CH230-127N19 (Children's Hospital Oakland
 Research Institute) complete sequence.

AC103453

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 199669)
 Muzny, D., Marle, D., Metzger, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, J., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Garcia, R., Garcia, A., Garner, T., Garza, M.,
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensdottir, L., Louise, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
 Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
 Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,
 Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okwuonu, G.,
 Olarunpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
 Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A.,
 Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E.,
 Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
 Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,
 Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S.,
 Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
 Sisson, I., Sitter, C.D., Smales, D., Sneed, A., Sodergren, E.,
 Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,
 Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
 Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D.,
 Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
 Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R.,
 Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, X.,
 Yen, J., Yoon, L., Yoon, Y., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
 Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R.,
 Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 199669)

Worley, K.C.

Direct Submission

Submitted (25-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 199669)

Worley, K.C.

Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 199669)

Worley, K.C.

Direct Submission

Submitted (07-JUN-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jun 7, 2003 this sequence version replaced gi:30521373.
 Sequencing is completed to a minimum standard of double strand
 coverage with a minimum of 2 clones and 2 reads with no ambiguities
 or 2 chemistries with a minimum of 2 clones and 3 reads with no
 ambiguities. If the sequence quality does not meet this standard,
 it will be indicated in the annotation.

FEATURES

source

Location/Qualifiers
 1. 199669
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /chromosome="13"
 /clone="CH230-127N19"

5394..5759

repeat_region

/rpt_family="Lx2_3"

5760..5861

repeat_region

/rpt_family="(CA)n"

5862..6447

repeat_region

/rpt_family="Lx2_3"

6522..6586

repeat_region

/rpt_family="(TG)n"

6611..6938

```

repeat_region      /rpt_family="L1M1"
6939..7037
/rpt_family="(TA)n"
repeat_region      complement(7044..7089)
/rpt_family="RMR1B"
7090..7169
/rpt_family="(TCTA)n"
repeat_region      complement(7190..7799)
/rpt_family="RMR1B"
repeat_region      complement(9049..10178)
/rpt_family="L1"
repeat_region      complement(10334..10982)
/rpt_family="L1"
repeat_region      10983..11076
/rpt_family="ID_Rn"
repeat_region      complement(11077..11432)
/rpt_family="L1"
repeat_region      11434..11887
/rpt_family="Lx6"
repeat_region      complement(12117..12370)
/rpt_family="RLTR28"
repeat_region      12371..13075
/rpt_family="L1VL4"
repeat_region      14687..14764
/rpt_family="ID4"
repeat_region      complement(17119..17216)
/rpt_family="ID_Rn"
repeat_region      complement(22339..22615)
/rpt_family="L1"
repeat_region      24851..24927
/rpt_family="(GA)n"
repeat_region      26001..26176
/rpt_family="Lx_3"
repeat_region      26183..26228
/rpt_family="L1"
repeat_region      complement(26236..26625)
/rpt_family="Lx_3"
repeat_region      26621..28264
/rpt_family="Lx2_3"
repeat_region      28430..28580
/rpt_family="L1"
repeat_region      complement(28667..29315)
/rpt_family="WTC-int"
repeat_region      complement(29316..29403)
/rpt_family="ID_Rn"
repeat_region      complement(29404..29418)
/rpt_family="WTC-int"
repeat_region      complement(29419..30342)
/rpt_family="L1"
repeat_region      complement(30346..30535)
/rpt_family="WTC-int"
repeat_region      30787..30932
/rpt_family="B3"
repeat_region      31007..31130
/rpt_family="L1"
repeat_region      31177..31306
/rpt_family="WTC-int"
repeat_region      31315..31418
/rpt_family="(TG)n"
repeat_region      31450..31883
/rpt_family="L1MB1"
repeat_region      31909..31988
/rpt_family="(GGGAGA)n"
repeat_region      32115..32163
/rpt_family="AT rich"
repeat_region      complement(32363..32408)
/rpt_family="MIR"
repeat_region      33092..33127
/rpt_family="(TC)n"
repeat_region      complement(33409..33631)
/rpt_family="L1MB8"
repeat_region      complement(33705..33796)
/rpt_family="L1MB8"

repeat_region      33877..34086
/rpt_family="B4A"
repeat_region      34995..35015
/rpt_family="AT rich"
repeat_region      35209..35296
/rpt_family="ID_Rn"
repeat_region      35666..35781
/rpt_family="RMR15"
repeat_region      36289..36443
/rpt_family="L1MA6"
repeat_region      36464..36605
/rpt_family="GA-rich"
repeat_region      36664..36875
/rpt_family="L1MA6"
repeat_region      36978..37520
/rpt_family="L1MA6"
repeat_region      37767..37998
/rpt_family="L1MB4"
repeat_region      38065..38215
/rpt_family="L1MB4"
repeat_region      38680..38849
/rpt_family="L1"
repeat_region      complement(39114..39170)
/rpt_family="MTD"
repeat_region      complement(39168..39315)

Query Match      68.8%; Score 28.2; DB 10; Length 199669;
Best Local Similarity 80.5%; Pred.No.1.3;
Matches 33; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY      1 GAATGAATACATAACAACACCTTACTGATCAGATCTACTTT 41
        |||||
Db      133709 GAATGACACATACGATCATCTTATCATATGCTATTTT 133749

RESULT 7
AC133259/c
LOCUS      AC133259      219071 bp      DNA      linear      HTG 20-NOV-2002
DEFINITION Rattus norvegicus clone CH230-318J10, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION  AC133259
VERSION     AC133259.2 GI:25139214
KEYWORDS    HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 219071)
AUTHORS    Muzny,D,Marie., Metzker,M,Lee., Abramson,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregregis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,

```

Lorensuhea, L., Loulseged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mamoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G. S., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelmech, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Prannkoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, D., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 219071)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (09-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 219071)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:22759157.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: KANT

Center clone name: CH230-318J10

----- Summary Statistics

Assembly program: Phrap; version 0.950329

Consensus quality: 194595 bases at least Q40

Consensus quality: 198491 bases at least Q30

Consensus quality: 200795 bases at least Q20

Estimated insert size: 202052; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 217199: contig of 217199 bp in length
* 217200 217299: gap of unknown length
* 217300 219071: contig of 1772 bp in length.

FEATURES
source

1..219071

/organism="Rattus norvegicus"

/mol_type="Genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-318J10"

1..1340

/note="wgs_contig"

3673..5446

/note="wgs_contig"

152568..153448

/note="clone_boundary"

clone_end:77

site:

end sequence:BZ142608"

212439..213833

/note="wgs_end_extension"

clone_end:77"

ORIGIN

Query Match 68.8%; Score 28.2; DB 2; Length 219071;

Best Local Similarity 80.5%; Pred. No. 1.3;

Matches 33; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GAATGAATCATACATACACACCTTATGTCAGATACACTTTT 41

DB 198951 GAATGACCATACATACATCATCTTTAATCATATGCTATTTT 198911

RESULT 8

AX346916

LOCUS

AX346916

DEFINITION

Sequence 1987 from Patent WO0200928.

ACCESSION

AX346916

VERSION

AX346916.1

GI:18494802

KEYWORDS

synthetic construct

artificial sequences.

ORGANISM

REFERENCE

1

AUTHORS

Olek, A., Piepenbrock, C. and Berlin, K.

TITLE

Diagnosis of diseases associated with the immune system

JOURNAL

Patent: WO 0200928-A 1987 03-JAN-2002;

Epigenomics AG (DE)

FEATURES

Location/Qualifiers

1..6182

/organism="synthetic construct"

/mol_type="unassigned DNA"

/db_xref="taxon:32630"

/note="Chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match

61.0%; Score 25; DB 6; Length 6182;

Best Local Similarity 75.6%; Pred. No. 32;

Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GAATGAATCATACATACACCTTATGTCAGATACACTTTT 41

DB 4872 GAATGAATCATATATATATTTTAGTAGATATATTTT 4912

RESULT 9

AC129295

LOCUS

AC129295

DEFINITION

Mus musculus BAC clone RP24-560M23 from chromosome 1, complete

121538 bp DNA linear ROD 27-NOV-2003

```

sequence.
AC129295
VERSION AC129295.4 GI:37651860
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Wang, C., Haakerson, W. and Haglund, K.
The sequence of Mus musculus BAC clone RP24-560M23
Unpublished (2001)
REFERENCE 2 (bases 1 to 121538)
AUTHORS Wilson, R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 121538)
AUTHORS McPherson, J. D. and Waterston, R. H.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 121538)
AUTHORS McPherson, J. D. and Waterston, R. H.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 121538)
AUTHORS Wilson, R. K.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 121538)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 11, 2003 this sequence version replaced gi:28867176.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Summary Statistics
-----
Center project name: M_BB0560M23
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.

```

```

FEATURES
    source
        1..121538
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /chromosome="1"
            /map="1"
            /clone="RP24-560M23"
            /clone_lib="RPCI-24"
        1..1753
            /rpt_family="L1"
        1801..1935
            /rpt_family="L1"
        5774..6000
            /rpt_family="L1"
        6013..6221
            /rpt_family="MaLR"
        6213..6690
            /rpt_family="MaLR"
        6830..7037
            /rpt_family="L1"
        7108..7293
            /rpt_family="B2"
        7347..7481
            /rpt_family="L1"
        7501..7597
            /rpt_family="MaLR"
        7663..8150
            /rpt_family="L1"
        8914..9125
            /rpt_family="MaLR"
        9341..9526
            /rpt_family="L1"
        10108..10198
            /rpt_family="L1"
        10274..10520
            /rpt_family="B4"
        10643..10724
            /rpt_family="L1"
        10974..11136
            /rpt_family="L1"
        11519..11625
            /rpt_family="ERVK"
        12864..13008
            /rpt_family="RMER15"
        13045..13270
            /rpt_family="RMER15"
        13801..13978
            /rpt_family="L1"
        14537..14672
            /rpt_family="L1"
        14909..15475
            /rpt_family="ERV1"
        15664..16263
            /rpt_family="L1"
        16711..16946
            /rpt_family="ERVK"
        17455..18017
            /rpt_family="L1"
        18013..18607
            /rpt_family="L1"
        18665..18990
            /rpt_family="ERVK"
        19007..19123
            /rpt_family="L1"
        19578..19724
            /rpt_family="B2"
        19909..19963
            /rpt_family="MaLR"
        19961..20188
            /rpt_family="MaLR"
        20307..20846
            /rpt_family="MaLR"
    repeat_region
        1801..1935
        5774..6000
        6013..6221
        6213..6690
        6830..7037
        7108..7293
        7347..7481
        7501..7597
        7663..8150
        8914..9125
        9341..9526
        10108..10198
        10274..10520
        10643..10724
        10974..11136
        11519..11625
        12864..13008
        13045..13270
        13801..13978
        14537..14672
        14909..15475
        15664..16263
        16711..16946
        17455..18017
        18013..18607
        18665..18990
        19007..19123
        19578..19724
        19909..19963
        19961..20188
        20307..20846

```

```

repeat_region 21301..22200
/rpt_family="L1"
repeat_region 29157..29428
/rpt_family="L1"
repeat_region 29859..31024
/rpt_family="L1"
repeat_region 31028..31144
/rpt_family="L1"
repeat_region 31170..32186
/rpt_family="L1"
repeat_region 32259..32415
/rpt_family="L1"
repeat_region 32867..33012
/rpt_family="Alu"
repeat_region 33079..33191
/rpt_family="L1"
repeat_region 33249..34124
/rpt_family="L1"
repeat_region 34600..35070
/rpt_family="L1"
repeat_region 35469..35546
/rpt_family="L1"
repeat_region 35347..35886
/rpt_family="MaLR"
repeat_region 35887..36282
/rpt_family="L1"
repeat_region 36487..36745
/rpt_family="L1"
repeat_region 37707..38105
/rpt_family="MaLR"
repeat_region 38106..38239
/rpt_family="B4"
repeat_region 38398..38463
/rpt_family="L2"
repeat_region 39337..39455
/rpt_family="Alu"
repeat_region 39511..39757
/rpt_family="MER1_type"
repeat_region 40057..40451
/rpt_family="MaLR"
repeat_region 40638..41045
/rpt_family="RMR17C"
repeat_region 43612..43654
/rpt_family="ERV1"
repeat_region 43676..43914
/rpt_family="ERVK"
repeat_region 44000..44432
/rpt_family="RMR17C"
repeat_region 45173..45288
/rpt_family="L1"
repeat_region 45732..46123
/rpt_family="MaLR"

Query Match 51.0%; Score 25; DB 10; Length 121538;
Best Local Similarity 75.8%; Pred. No. 25;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GAATGAACATACATACACACACACCTTAGTCAGATACACTTT 41
Db 115259 GAATGAACATACATACATACATCTTTATACAGGCTATTT 115299

RESULT 10
AC134414/C
LOCUS AC134414 200844 bp DNA linear HTG 26-FEB-2003
DEFINITION Mus musculus clone RP24-70H12, WORKING DRAFT SEQUENCE, 13 unordered
pieces.
ACCESSION AC134414
VERSION AC134414.3 GI:28565730
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 200844)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gird,S., Graham,L., Grand-pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Naylor,J.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 200844)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K., Faro,S.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Gardyna,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramaamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-FEB-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 26, 2003 this sequence version replaced gi:28275018.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26735
Center clone name: 70_H12
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 197943 bases at least Q40
Consensus quality: 199015 bases at least Q30
Consensus quality: 199440 bases at least Q20
Insert size: 194000; agarose-ep
Insert size: 199644; sum-of-contigs
Quality coverage: 9.2 in Q20 bases; agarose-fp

```

Quality coverage: 8.9 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 54386: contig of 54386 bp in length
* 54387 54486: gap of 100 bp
* 54487 55846: contig of 1360 bp in length
* 55847 55946: gap of 100 bp
* 55947 59197: contig of 3251 bp in length
* 59198 59297: gap of 100 bp
* 59298 64542: contig of 5245 bp in length
* 64543 64642: gap of 100 bp
* 64643 69113: contig of 4471 bp in length
* 69114 69213: gap of 100 bp
* 69214 72896: contig of 3683 bp in length
* 72897 72996: gap of 100 bp
* 72997 78283: contig of 5287 bp in length
* 78284 78384: gap of 100 bp
* 78384 90557: contig of 12174 bp in length
* 90558 90657: gap of 100 bp
* 90658 10316: contig of 10659 bp in length
* 10317 10416: gap of 100 bp
* 10417 120870: contig of 19454 bp in length
* 120871 120970: gap of 100 bp
* 120971 145380: contig of 24410 bp in length
* 145381 145480: gap of 100 bp
* 145481 174041: contig of 28561 bp in length
* 174042 174141: gap of 100 bp
* 174142 200844: contig of 26703 bp in length.

```

FEATURES

```

Source
1. 200844
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /clone="RP24-70H12"
   /clone_lib="RPCI-24 Male Mouse BAC"
misc_feature
1. 54386
   /note="assembly_fragment"
   /vector_side:left
misc_feature
54487..55846
   /note="assembly_fragment"
misc_feature
55947..59197
   /note="assembly_fragment"
misc_feature
59298..64542
   /note="assembly_fragment"
misc_feature
64643..69113
   /note="assembly_fragment"
misc_feature
69214..72896
   /note="assembly_fragment"
misc_feature
72997..78283
   /note="assembly_fragment"
misc_feature
78384..90557
   /note="assembly_fragment"
misc_feature
90658..101316
   /note="assembly_fragment"
misc_feature
101417..120870
   /note="assembly_fragment"
misc_feature
120971..145380
   /note="assembly_fragment"
misc_feature
145481..174041
   /note="assembly_fragment"
misc_feature
174142..200844
   /note="assembly_fragment"
   /vector_side:right

```

ORIGIN

```

Query Match      61.0%; Score 25; DB 2; Length 200844;
Best Local Similarity 75.6%; Pred. No. 24;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Oy 1 GAATGACTACATTAACACACCTTACTGACTGACTACTCTTT 41
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60430 GAATACACACAGAAATTAAGACCTTGCTCAATATTACTTT 60390

RESULT 11
AC110569/c
LOCUS AC110569 213457 bp DNA linear HTG 21-OCT-2002
DEFINITION Mus musculus clone RP23-309J17, WORKING DRAFT SEQUENCE, 7 ordered
pieces.
ACCESSION AC110569
VERSION AC110569.3 GI:24182091
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-309J17
Unpublished
2 (bases 1 to 213457)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Gardyna,S., Gord,S., Goyette,M., Graham,I., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Labrecque,K., Lamaze,R.,
Landers,T., Lehotzky,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McSwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rhee,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 213457)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,I., Grand-Pierre,N., Hagel,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrum,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupbach,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome

```


COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Oct 21, 2002 this sequence version replaced gi:20503222.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIEB
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L22620
 Center clone name: 309 J.17
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a, 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 212308 bases at least Q40
 Consensus quality: 212725 bases at least Q30
 Consensus quality: 212828 bases at least Q20
 Insert size: 210000; agarose-fp
 Insert size: 212857; sum-of-contigs
 Quality coverage: 11.3 in Q20 bases; agarose-fp
 Quality coverage: 11.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1
 * 65944: contig of 65944 bp in length
 * 65945: gap of 100 bp
 * 66045: contig of 3638 bp in length
 * 69683: gap of 100 bp
 * 69783: contig of 3875 bp in length
 * 73558: gap of 100 bp
 * 73758: contig of 60623 bp in length
 * 134381: gap of 100 bp
 * 134481: contig of 28011 bp in length
 * 162491: gap of 100 bp
 * 162492: contig of 38227 bp in length
 * 200918: gap of 100 bp
 * 200919: contig of 12539 bp in length.

FEATURES

source
 1. .213457
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP23-309017"
 /clone_lib="RPCI-23 Female Mouse BAC"
 1. .65944
 /notes="assembly_fragment"
 clone_end:SP6
 vector_side:left
 66045..69682
 /notes="assembly_fragment"
 69783..73657
 /notes="assembly_fragment"
 73758..134380
 /notes="assembly_fragment"
 134481..162491
 /notes="assembly_fragment"
 162592..200818
 /notes="assembly_fragment"
 200919..213457
 /notes="assembly_fragment"
 clone_end:T7
 vector_side:right

ORIGIN

Query Match 61.0%; Score 25; DB 2; Length 213457;
 Best Local Similarity 75.8%; Pred. No. 24;
 Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 1 GAATGAACCTACATACACACACCTTAGTCAGATACACTTT 41
 Db 58385 GAATAAACACAGATTAGACCTTGTCATATATTACTTT 58345
 RESULT 12
 AL672023/c
 LOCUS
 DEFINITION
 Mouse DNA sequence from clone RP23-339019 on chromosome X, complete
 sequence.
 AL672023
 VERSION
 AL672023.7 GI:21621646
 KEYWORDS
 HTG.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Tracey,A.
 Direct Submission
 Submitted (26-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Jun 27, 2002 this sequence version replaced gi:21531223.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw:
 SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-339019 is
 from the RPCI-23 Mouse PAC Library
 constructed by the group of Pieter de Jong.
 For further details see http://www.chori.org/bacpac/home.htm
 VECTOR: PBACE3.6. Location/Qualifiers

FEATURES

source
 1. .160615
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-339019"
 /clone_lib="RPCI-23"

ORIGIN

Query Match 60.0%; Score 24.6; DB 10; Length 160615;
 Best Local Similarity 76.9%; Pred. No. 35;
 Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 1 GAATGAACCTACATACACACCTTAGTCAGATACACTTT 39
 Db 58385 GAATAAACACAGATTAGACCTTGTCATATATTACTTT 58345

Db 21328 GACTGCCACATACACACCACTCAGACACTATT 21290

RESULT 13
HSM804835 3473 bp mRNA linear PRI 13-MAY-2003
LOCUS HSM804835 3473 bp mRNA linear PRI 13-MAY-2003
DEFINITION Homo sapiens mRNA; cDNA DKFZp686G0837 (from clone DKFZp686G0837).
ACCESSION AL833522
VERSION AL833522.1 GI:21734166
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3473)
AUTHORS Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
Fobo,G., Han,M. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp686G0837) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at <http://mips.gsf.de/proj/cDNA/>.
FEATURES
source
1..3473
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686G0837"
/tissue_type="human testis"
/clone_lib="686 (synonym: hlcc3). Vector pSport1_sfi; host
DH10B; sites SfiIA + SfiIB"
/dev_stage="adult"
polyA_signal 3435..3440
polyA_site 3452

ORIGIN
Query Match 58.5%; Score 24; DB 9; Length 3473;
Best Local Similarity 75.0%; Pred. No. 83;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Qy 1 GAATGAACATACATACACCACTTACTGATGACTACTTT 40
Db 1143 GCATATACATAAATAGTTCAGCTAACTCAGCTACTACTT 1182

RESULT 14
AX251133/c 5276 bp DNA linear PAT 05-OCT-2001
LOCUS AX251133/c 5276 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 101 from Patent WO0168912.
ACCESSION AX251133
VERSION AX251133.1 GI:15984556
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and
oncogenes
JOURNAL Patent: WO 0168912-A 101 20-SEP-2001;
Epigenomics AG (DE)
FEATURES
source
1..5276
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"

ORIGIN
/note="chemically treated genomic DNA (Homo sapiens)"
Query Match 58.5%; Score 24; DB 6; Length 5276;
Best Local Similarity 75.0%; Pred. No. 80;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Qy 2 AATGAACATACATACACCACTTACTGATGACTACTTT 41
Db 544 AATTACCTACATACCAAACTCATAAATCTCAATTTCTACATT 505

RESULT 15
AX345728/c 5276 bp DNA linear PAT 01-FEB-2002
LOCUS AX345728/c 5276 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 799 from Patent WO0200928.
ACCESSION AX345728
VERSION AX345728.1 GI:18493614
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 799 03-JAN-2002;
Epigenomics AG (DE)
FEATURES
source
1..5276
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN
Query Match 58.5%; Score 24; DB 6; Length 5276;
Best Local Similarity 75.0%; Pred. No. 80;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Qy 2 AATGAACATACATACCAACCACTTACTGATGACTACTTT 41
Db 544 AATTACCTACATACCAAACTCATAAATCTCAATTTCTACATT 505

Search completed: March 24, 2004, 23:01:31
Job time : 911.607 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 18:56:32 ; Search time 214.674 Seconds
(without alignments)
811.351 Million cell updates/sec

Title: US-09-889-611a-60_COPY_2_42

Perfect score: 41

Sequence: 1 gaatgaactacatacaacc.....ccttagtcagatactacttt 41

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001as.*

5: Geneseq2001bs.*

6: Geneseq2002s.*

7: Geneseq2003as.*

8: Geneseq2003bs.*

9: Geneseq2003cs.*

10: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	1431	3	AAA71435 Human meg
2	40	97.6	128	3	AAA71434 Human meg
3	33.6	82.0	6182	6	ABL34015 Human imm
4	30	73.2	30	3	AAA71449 Human meg
5	30	73.2	30	3	AAA71448 Human meg
6	25	61.0	25	3	AAA71453 Human meg
7	25	61.0	6182	6	ABL34014 Human imm
8	24	58.5	5276	4	AA846379 Tumour su
9	24	58.5	5276	6	ABL32826 Human imm
10	23	56.1	23	3	AAA71452 Human meg
11	22.4	54.6	1380	9	ADC92907 E. faeciu
12	22.4	54.6	5204	6	ABL32899 Human imm
13	22.4	54.6	11052	6	ABK39985 Human che
14	22	53.7	6503	6	ABL32720 Human imm
15	21.8	53.2	2222	6	ABK34986 Human cDN
16	21.8	53.2	3561	4	ABL13848 Drosophil
17	21.8	53.2	6568	6	AA894755 Human DNA
18	21.6	52.7	1265	5	AA668177 Human lun
19	21.6	52.7	1265	6	ABK38088 cDNA enco
20	21.6	52.7	1265	7	ACAL0417 Human lun
21	21.6	52.7	1265	7	ABX99368 Lung canc
22	21.6	52.7	1265	10	ABE72151 Human lun
23	21.6	52.7	1350	5	AA872599 DNA enco

ALIGNMENTS

RESULT 1

AAA71435

ID AAA71435 standard; DNA; 1431 BP.

AC AAA71435;

XX

DT 01-DEC-2000 (first entry)

XX

DE Human megsin promoter fragment DNA.

XX

KW Promoter; megsin; human; protein isolation; screening. ss.

XX

OS Homo sapiens.

XX

PN WO200043528-A1.

XX

PD 27-JUL-2000.

XX

PF 25-JAN-2000; 2000WO-JP000350.

XX

PR 25-JAN-1999; 99JP-00015667.

XX

PA (KURO/) KUROKAWA K.

XX

PI (MIYA/) MIYATA T.

XX

XX Miyata T;

XX

DR WPI; 2000-543257/49.

XX

PT DNA for promoter region of megsin useful for screening proteins.

XX

PS Disclosure; Fig 2; 45pp; Japanese.

XX

CC This invention describes a novel DNA sequence (I) representing a promoter region having part or all of a specific base sequence. The invention also describes (1) a vector containing (I); (2) a cell transformed by the above vector; and (3) protein produced using (I). (I) is useful for screening and isolating proteins (especially transcription factors). This sequence represents a fragment of the human megsin promoter which is described in the method of the invention

XX

SQ Sequence 1431 BP; 466 A; 266 C; 303 G; 394 T; 0 U; 2 Other;

Query Match 100.0%; Score 41; DB 3; Length 1431;

Best Local Similarity 100.0%; Pred. No. 2.8e-06;

Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATGAAGTACATACATACCAACACCTTAGTCAGATACTACTTT 41
 DB 1303 GAATGAAGTACATACATACCAACACCTTAGTCAGATACTACTTT 1343

RESULT 2
 AAA71434
 ID AAA71434 standard; DNA; 128 BP.
 AC AAA71434;
 XX
 XX 01-DEC-2000 (first entry)
 DT
 DE Human meglin promoter fragment DNA.
 XX
 XX Promoter; meglin; human; protein isolation; screening. ss.
 KW Homo sapiens.
 OS
 XX WO200043528-A1.
 PN
 XX 27-JUL-2000.
 PD
 XX 25-JAN-2000; 2000WO-JP000350.
 PF
 XX 25-JAN-1999; 99JP-00015667.
 PR
 XX (KURO/) KUROKAWA K.
 PA (MIYA/) MIYATA T.
 XX
 XX Miyata T;
 PI
 XX WPI; 2000-543257/49.
 DR
 XX
 XX DNA for promoter region of meglin useful for screening proteins.
 PT
 XX
 XX Claim 1; Page 32; 45pp; Japanese.

CC This invention describes a novel DNA sequence (I) representing a promoter region having part or all of a specific base sequence. The invention also describes (1) a vector containing (I); (2) a cell transformed by the above vector; and (3) protein produced using (I). (I) is useful for screening and isolating proteins (especially transcription factors). This sequence represents the human meglin promoter which is described in the method of the invention

XX
 SQ Sequence 128 BP; 44 A; 29 C; 24 G; 29 T; 0 U; 2 Other;
 Query Match 97.6%; Score 40; DB 3; Length 128;
 Best Local Similarity 100.0%; Pred. No. 4.6e-06;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AATGAAGTACATACATACCAACACCTTAGTCAGATACTACTTT 41
 DB 1 AATGAAGTACATACATACCAACACCTTAGTCAGATACTACTTT 40

RESULT 3
 ABL34015/c
 ID ABL34015 standard; DNA; 6182 BP.
 XX
 AC ABL34015;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 XX Human immune system associated gene SEQ ID NO: 1988.
 DE
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 DS.
 XX Homo sapiens.
 OS
 XX WO200200928-A2.
 PN
 XX 03-JAN-2002.
 PD
 XX 02-JUL-2001; 2001WO-EP007537.
 PF
 XX 30-JUN-2000; 2000DE-01032529.
 PR
 XX 01-SEP-2000; 2000DE-01043826.
 XX
 XX (BPIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 DR
 XX
 XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX
 XX Claim 1; SEQ ID NO 1988; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention

XX
 SQ Sequence 6182 BP; 1934 A; 37 C; 1059 G; 3152 T; 0 U; 0 Other;
 Query Match 82.0%; Score 33.6; DB 6; Length 6182;
 Best Local Similarity 90.0%; Pred. No. 0.0029;
 Matches 36; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AATGAAGTACATACATACCAACACCTTAGTCAGATACTACTTT 41
 DB 1310 AATGAAGTACATACATACCAACACCTTAGTCAGATACTACTTT 1271

RESULT 4
 AAA71449
 ID AAA71449 standard; DNA; 30 BP.
 XX
 AC AAA71449;
 XX
 XX 01-DEC-2000 (first entry)
 DT
 XX Human meglin promoter PCR primer SEQ ID NO: 16.
 DE
 KW Promoter; meglin; human; protein isolation; screening. PCR primer; ss.
 XX Homo sapiens.
 OS
 XX WO200043528-A1.
 PN
 XX 27-JUL-2000.
 PD
 XX 25-JAN-2000; 2000WO-JP000350.
 PF
 XX 25-JAN-1999; 99JP-00015667.
 PR
 XX (KURO/) KUROKAWA K.
 PA (MIYA/) MIYATA T.
 XX
 XX Miyata T;

```

XX DR WPI; 2000-543257/49.
XX PT DNA for promoter region of megins useful for screening proteins.
XX PS Example 5; Page 40; 45pp; Japanese.
XX CC This invention describes a novel DNA sequence (I) representing a promoter
CC region having part or all of a specific base sequence. The invention also
CC describes (1) a vector containing (I); (2) a cell transformed by the
CC above vector; and (3) protein produced using (I). (I) is useful for
CC screening and isolating proteins (especially transcription factors).
CC AAA71434-A71469 represent PCR primers used in the method described in the
CC invention
XX SQ Sequence 30 BP; 12 A; 9 C; 2 G; 7 T; 0 U; 0 Other;

Query Match 73.2%; Score 30; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.031;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 TACATAACCAACCACTTAGTCAGATACTAC 38
Db 1 TACATAACCAACCACTTAGTCAGATACTAC 30

RESULT 5
AAA71448
ID AAA71448 standard; DNA; 30 BP.
XX AC AAA71448;
XX DT 01-DEC-2000 (first entry)
XX DE Human megins promoter PCR primer SEQ ID NO: 15.
XX KW Promoter; megins; human; protein isolation; screening. PCR primer; ss.
XX OS Homo sapiens.
XX PN WO200043528-A1.
XX PD 27-JUL-2000.
XX PF 25-JAN-2000; 2000WO-JP000350.
XX PR 25-JAN-1999; 99JP-00015667.
XX PA (KURO/) KUROKAWA K.
XX PA (MIYA/) MIYATA T.
XX PI Miyata T;
XX DR WPI; 2000-543257/49.
XX PT DNA for promoter region of megins useful for screening proteins.
XX PS Example 5; Page 40; 45pp; Japanese.
XX CC This invention describes a novel DNA sequence (I) representing a promoter
CC region having part or all of a specific base sequence. The invention also
CC describes (1) a vector containing (I); (2) a cell transformed by the
CC above vector; and (3) protein produced using (I). (I) is useful for
CC screening and isolating proteins (especially transcription factors).
CC AAA71434-A71469 represent PCR primers used in the method described in the
CC invention
XX SQ Sequence 30 BP; 13 A; 8 C; 3 G; 6 T; 0 U; 0 Other;

Query Match 73.2%; Score 30; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.031;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2 AATGAACCTACATAACCAACCACTTAGTCAG 31
Db 1 AATGAACCTACATAACCAACCACTTAGTCAG 30

RESULT 6
AAA71453
ID AAA71453 standard; DNA; 25 BP.
XX AC AAA71453;
XX DT 01-DEC-2000 (first entry)
XX DE Human megins promoter PCR primer SEQ ID NO: 20.
XX KW Promoter; megins; human; protein isolation; screening. PCR primer; ss.
XX OS Homo sapiens.
XX PN WO200043528-A1.
XX PD 27-JUL-2000.
XX PF 25-JAN-2000; 2000WO-JP000350.
XX PR 25-JAN-1999; 99JP-00015667.
XX PA (KURO/) KUROKAWA K.
XX PA (MIYA/) MIYATA T.
XX PI Miyata T;
XX DR WPI; 2000-543257/49.
XX PT DNA for promoter region of megins useful for screening proteins.
XX PS Example 4; Page 42; 45pp; Japanese.
XX CC This invention describes a novel DNA sequence (I) representing a promoter
CC region having part or all of a specific base sequence. The invention also
CC describes (1) a vector containing (I); (2) a cell transformed by the
CC above vector; and (3) protein produced using (I). (I) is useful for
CC screening and isolating proteins (especially transcription factors).
CC AAA71434-A71469 represent PCR primers used in the method described in the
CC invention
XX SQ Sequence 25 BP; 8 A; 7 C; 2 G; 8 T; 0 U; 0 Other;

Query Match 61.0%; Score 25; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 AACCACTTAGTCAGATACTACTTT 41
Db 1 AACCACTTAGTCAGATACTACTTT 25

RESULT 7
ABL34014
ID ABL34014 standard; DNA; 6182 BP.
XX AC ABL34014;
XX DT 26-MAR-2002 (first entry)
XX DE Human immune system associated gene SEQ ID NO: 1987.
XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

```

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX Homo sapiens.
XX
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007537.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX
XX Claim 1; SEQ ID NO 1987; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention
XX
XX Sequence 6182 BP; 1929 A; 37 C; 1260 G; 2956 T; 0 U; 0 Other;
XX
Query Match 61.0%; Score 25; DB 6; Length 6182;
Best Local Similarity 75.6%; Pred. No. 6.6;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
XX
QY 1 GAATGAATCATATACACACACCTTAGTCAGATACACTTTT 41
DB 4872 GAATGAATATATATATATTTTATTAGTTAGTATATTATTT 4912
XX
RESULT 8
AAS46379/C
ID AAS46379 standard; DNA; 5276 BP.
XX
XX AAS46379;
XX
XX 18-DEC-2001 (first entry)
XX
XX Tumour suppressor gene derived chemically modified sequence #101.
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
XX tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX cytosine methylation; ds.
XX
XX Homo sapiens.
XX
XX WO200168912-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP002955.
XX
XX 15-MAR-2000; 2000DE-01013847.
XX
XX 06-APR-2000; 2000DE-01019058.
XX
XX 07-APR-2000; 2000DE-01019173.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX

XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2001-602752/68.
XX
XX Fragments of chemically modified genes associated with tumor suppressor
XX genes and oncogenes, useful in designing primers and probes for analyzing
XX diseases associated with cytosine methylation state e.g. cancer.
XX
XX Claim 1; SEQ ID NO 101; 27pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
XX bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX bisulphite, of genes associated with tumour suppression and oncogenes
XX having a sequence taken from 536 (actually 533 since numbers 408, 458 and
XX 500 are missing from the sequence listing) sequences (Ss) and sequences
XX complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
XX oligomer (PNA) of at least 9 nucleotides and may form part of a set of
XX probes for detecting the cytosine methylation state and/or single
XX nucleotide polymorphisms and also to be used in an array for analysing
XX diseases associated with CpG dinucleotides e.g. cancers and tumours. The
XX probes can also be used in a method for ascertaining genetic and/or
XX epigenetic parameters for the diagnosis and/or therapy of existing
XX diseases or the predisposition to specific diseases, by analysing
XX cytosine methylations. The parameters may be compared to another set of
XX genetic and/or epigenetic parameters, the differences serving as basis
XX for diagnosis and/or prognosis events which are disadvantageous to
XX patients. The present sequence is one of the 533 genomic sequences
XX derived from tumour suppressor genes and oncogenes. Note: The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 5276 BP; 1546 A; 88 C; 1102 G; 2540 T; 0 U; 0 Other;
XX
Query Match 58.5%; Score 24; DB 4; Length 5276;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
XX
QY 2 AATGACTACATACACACACCTTAGTCAGATACACTTTT 41
DB 544 AATTACCTACATACACAACTCACTAACTCAAAATTCATTT 505
XX
RESULT 9
ABL32826/C
ID ABL32826 standard; DNA; 5276 BP.
XX
XX ABL32826;
XX
XX 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 799.
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; anti-anaemic; cytostatic; nootropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.
XX
XX Homo sapiens.
XX
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007537.
XX

PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 PA (EPIG-) EPIGENOMICS AG.
 XX Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 DR Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX Claim 1; SEQ ID NO 799; 32pp + Sequence Listing; German.
 PS The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and,
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX
 SQ Sequence 5276 BP; 1546 A; 88 C; 1102 G; 2540 T; 0 U; 0 Other;
 Query Match 58.5%; Score 24; DB 6; Length 5276;
 Best Local Similarity 75.0%; Pred. No. 16;
 Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 2 AATGAACTACATACACACCTTAGTCAGATCTACTTT 41
 DB 544 AATTACCTACATACAAACTCAATCAATCAATTCATTT 505
 RESULT 10
 AAA71452
 ID AAA71452 standard; DNA; 23 BP.
 AC AAA71452;
 XX 01-DEC-2000 (first entry)
 DT Human meggin promoter PCR primer SEQ ID NO: 19.
 DE Promoter; meggin; human; protein isolation; screening. PCR primer; ss.
 XX Homo sapiens.
 OS WO200043528-A1.
 PN 27-JUL-2000.
 PD 25-JAN-2000; 2000WO-JP000350.
 PF 25-JAN-1999; 99JP-00015667.
 PR (KURO/) KUROKAWA K.
 PA (MIYA/) MIYATA T.
 XX Miyata T;
 XX WPI; 2000-543257/49.
 DR DNA for promoter region of meggin useful for screening proteins.
 XX Example 4; Page 42; 45pp; Japanese.
 PS This invention describes a novel DNA sequence (I) representing a promoter
 CC region having part or all of a specific base sequence. The invention also
 CC describes (1) a vector containing (I); (2) a cell transformed by the
 CC above vector; and (3) protein produced using (I). (I) is useful for
 CC screening and isolating proteins (especially transcription factors).
 CC

CC AAA71434-A71469 represent PCR primers used in the method described in the
 CC invention
 XX Sequence 23 BP; 11 A; 7 C; 2 G; 3 T; 0 U; 0 Other;
 SQ Query Match 56.1%; Score 23; DB 3; Length 23;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAATGAACATACATAAACAACACACC 23
 DB 1 GAATGAACATACATAAACAACACACC 23
 RESULT 11
 ADC92907/c
 ID ADC92907 standard; DNA; 1380 BP.
 XX AC ADC92907;
 XX 01-JAN-2004 (first entry)
 DT E. faecium DNA sequence SEQ ID 2534.
 XX ds; gene; urinary tract infection; bacteraemia; endocarditis; wound;
 XX abdominal-pelvic infection.
 OS Enterococcus faecium.
 PN US6583275-B1.
 XX 24-JUN-2003.
 XX 30-JUN-1998; 98US-00107532.
 XX 02-JUL-1997; 97US-0051571P.
 PR 14-MAY-1998; 98US-0085598P.
 XX (GENO-) GENOME THERAPEUTICS CORP.
 XX Doucette-Stamm LA, Bush D;
 XX WPI; 2003-799836/75.
 DR P-PSDB; ADC96561.
 XX New isolated nucleic acid derived from Enterococcus faecium encoding an
 PT Enterococcus faecium polypeptide useful for detection, prevention and
 PT treatment of a pathological condition resulting from a bacterial
 PT infection.
 XX Example 1; SEQ ID NO 2534; 243pp; English.
 XX The invention relates to an isolated nucleic acid derived from
 CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
 CC one of 10 fully defined sequences given in the (or comprising 40
 CC sequential nucleotides chosen from any of the nucleic acids, its
 CC complement or sequences hybridising to it). Also included are a
 CC recombinant vector comprising the nucleic acid operably linked to
 CC transcription regulatory element, a cell comprising the vector and a
 CC single-stranded probe comprising the nucleic acid. The nucleic acids are
 CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
 CC The nucleic acids is useful for diagnosing pathological conditions
 CC resulting from E. faecium bacterial infection (e.g. urinary tract
 CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
 CC infection) and for screening drugs such as agonists and antagonists. The
 CC nucleic acid is useful for recombinant production of Candida albicans -
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions
 CC and vaccines containing the nucleic acid are useful for preventing or
 CC treating Enterococcus faecium infections. The present sequence represents
 CC one if the disclosed E. faecium nucleic acids.
 XX Sequence 1380 BP; 475 A; 213 C; 306 G; 386 T; 0 U; 0 Other;
 SQ

Query Match 54.6%; Score 22.4; DB 9; Length 1380;
 Best Local Similarity 72.5%; Pred. No. 55;
 Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 GAATGACTACATACCAACACCTTAGTCAGATACACTTT 40
 DB 1019 GATTGACCTACATACCAACCAACCAACCAACGATTTCTCTT 980

RESULT 12
 ABL32899/c

ID ABL32899 standard; DNA; 5204 BP.
 AC ABL32899;
 XX
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 872.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosine methylation; nontropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.
 OS Homo sapiens.
 XX
 XX WO200200928-A2.
 PN
 XX
 DT 03-JAN-2002.
 XX
 PD
 XX
 PF 02-JUL-2001; 2001WO-EP007537.
 XX
 XX 30-JUN-2000; 2000DE-01032529.
 PR
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2002-130909/17.
 DR
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX
 XX Claim 1; SEQ ID NO 872; 32pp + Sequence Listing; German.
 PS
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX
 XX Sequence 5204 BP; 1302 A; 79 C; 1119 G; 2704 T; 0 U; 0 Other;

Query Match 54.6%; Score 22.4; DB 6; Length 5204;
 Best Local Similarity 72.5%; Pred. No. 68;
 Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 AATGAACATACATACCAACACCTTAGTCAGATACACTTT 41
 DB 1180 AATGAACATACATACCAACCAACCAACCAACATTAACACTTT 1141

RESULT 13
 ABK39985/c

ID ABK39985 standard; DNA; 11052 BP.
 AC ABK39985;
 XX
 XX
 DT 21-MAY-2002 (first entry)
 XX
 DE Human chemically pretreated gene sequence #34 strand 1.
 XX
 KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
 KW cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
 KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
 XX
 XX Homo sapiens.
 OS
 XX WO200202806-A2.
 PN
 XX
 DT 10-JAN-2002.
 XX
 PD
 XX
 PF 29-JUN-2001; 2001WO-EP007470.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 PR
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2002-154757/20.
 DR
 XX
 XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
 XX useful for detecting cytosine methylation state of genes associated with
 XX pharmacogenomics and for therapy of diseases e.g. cancer.
 XX
 XX Claim 1; SEQ ID NO 67; 24pp; English.

The invention relates to a nucleic acid comprising a sequence at least 18
 bases in length of a segment of the chemically pretreated DNA of genes
 associated with pharmacogenomics according to one of the sequences of the
 genes ALDH6 (NM 000693), CYP11A (NM 000781), CYP11B1 (NM 000497), CYP3A3
 (NM 000776 and NM 017460), DPYD (NM 000110), EPHX2 (NM 001979), OCLN
 (NM 002538), TXNRD1 (NM 003330), UGT8 (NM 003360), MRP (NM 004996),
 NM 019900, NM 019901, NM 019902, NM 019862, NM 019898, NM 019899) and
 their complementary sequences, or a sequence (S1) chosen from 87
 sequences and their complements. The chemical pretreatment is bisulphite
 treatment to convert cytosines (but not methyl-cytosines) into uracils.
 Also included are an oligomer (II) in particular an oligonucleotide or a
 peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
 base sequence having a length of 9 nucleotides which hybridises to or is
 identical to a chemically pretreated DNA of genes associated with
 pharmacogenomics and their complements, arranged in an array for
 analysing diseases associated with the methylation state (CpG) and/or
 detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
 oligomers may also be used as PCR primers. The set of 87 nucleic acids
 and their complements is useful for diagnosis and therapy of solid
 tumours and cancer. The present sequence represents one the 87 DNA
 sequences or its complement. Note: The sequence data for this patent did
 not form part of the printed specification, but was obtained in
 electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences

XX
 XX Sequence 11052 BP; 2796 A; 167 C; 2778 G; 5311 T; 0 U; 0 Other;

Query Match 54.6%; Score 22.4; DB 6; Length 11052;
 Best Local Similarity 72.5%; Pred. No. 76;
 Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 AATGAACATACATACCAACACCTTAGTCAGATACACTTT 41
 DB 10756 AATTAACATACATCAACACCACTTAATTAACACTTT 10717

RESULT 14
 ABL32720/c

ID ABL32720 standard; DNA; 6503 BP.
 XX AC ABL32720;
 XX DT 26-MAR-2002 (first entry)
 XX DE Human immune system associated gene SEQ ID NO: 693.
 XX DE Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antileukemia; anti-HIV; anticonvulsant; ophthalmologic;
 KW antineoplastic; antiarthritic; antidiabetic; antipsoriatic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.
 XX OS Homo sapiens.
 XX PN WO200200928-A2.
 XX PD 03-JAN-2002.
 XX PF 02-JUL-2001; 2001WO-BP007537.
 XX PR 30-JUN-2000; 2000DE-01032529.
 XX PR 01-SEP-2000; 2000DE-01043826.
 XX PA (EPIG-) EPIGENOMICS AG.
 XX PI Olek A, Piepenbrock C, Berlin K;
 XX DR WPI; 2002-130909/17.
 XX PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX PS Claim 1; SEQ ID NO 693; 32pp + Sequence Listing; German.
 XX CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX SQ Sequence 6503 BP; 1561 A; 85 C; 1446 G; 3411 T; 0 U; 0 Other;
 Query Match 53.7%; Score 22; DB 6; Length 6503;
 Best Local Similarity 73.7%; Pred. No. 1e+02;
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 2 AATGAACATACATACCAACCACTTAGTCAGATCTACT 39
 DB 509 AATAAACAACATATAAATACCAATCGTCCTATATTACT 472
 RESULT 15
 ABK34986/c
 ID ABK34986 standard; cDNA; 2222 BP.
 XX AC ABK34986;
 XX DT 08-MAY-2002 (first entry)
 XX DE Human cDNA encoding secreted protein #124.
 KW Human; secreted protein; gene; ss; nutritional supplement; haemophilia;
 KW viral infection; bacterial infection; fungal infection; diabetes; asthma;
 KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;

KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
 KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
 KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;
 KW tissue regeneration; wound healing; burn; haematopoiesis;
 KW myeloid cell deficiency; lymphoid cell deficiency.
 XX OS Homo sapiens.
 XX PN WO200177288-A2.
 XX PD 18-OCT-2001.
 XX PF 29-MAR-2001; 2001WO-US010224.
 XX PR 06-APR-2000; 2000US-0195582P.
 XX PA (GEMV) GENETICS INST INC.
 XX PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
 XX PI Gulukota K, Graham JR;
 XX DR WPI; 2002-179321/23.
 XX PF Five hundred and ninety two polynucleotides derived from a variety of
 PT human tissue sources which encode secreted proteins, useful for treating
 PT immune deficiencies and disorders such as autoimmune disorders.
 XX PS Claim 1; Page 130-131; 372pp; English.
 XX CC The invention relates to 592 polynucleotides which have been derived from
 CC a variety of human tissue sources and which encode novel secreted
 CC proteins. The polynucleotides can be used as probes for the
 CC identification and isolation of full length cDNA and genomic DNA. The
 CC polynucleotides and proteins can also be used as nutritional supplements.
 CC The proteins are useful in the treatment of various immune deficiencies
 CC and disorders such as viral infections, bacterial infections, fungal
 CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
 CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
 CC and conditions (e.g. asthma). They are also useful for treating
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
 CC useful for tissue regeneration, for wound healing and in the treatment of
 CC burns, incisions and ulcers. The proteins are also useful for regulating
 CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.
 CC Sequences ABK34863-ABK35454 represent polynucleotides of the invention
 XX SQ Sequence 2222 BP; 573 A; 581 C; 515 G; 553 T; 0 U; 0 Other;
 Query Match 53.2%; Score 21.8; DB 6; Length 2222;
 Best Local Similarity 70.7%; Pred. No. 1e+02;
 Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
 QY 1 GAATGAACATACATACCAACCACTTAGTCAGATCTACTTT 41
 DB 2037 GAATGAACATACACCAACCAAGGATCTTTGTTCAATACAAATGT 1997
 Search completed: March 24, 2004, 22:28:26
 Job time : 217.674 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 24, 2004, 21:57:32 ; Search time 1996.1 Seconds
(without alignments)
613.370 Million cell updates/sec

Title: US-09-889-611a-60_COPY_2_42

Perfect score: 41
Sequence: 1 gaatgaactacatacaacc.....ccttagtcagatactacttt 41

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hcc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gaa_hum:*
18: em_gaa_inv:*
19: em_gaa_pin:*
20: em_gaa_vrt:*
21: em_gaa_fun:*
22: em_gaa_mam:*
23: em_gaa_mus:*
24: em_gaa_pro:*
25: em_gaa_rtd:*
26: em_gaa_pug:*
27: em_gaa_vrt:*
28: gb_gaa1:*
29: gb_gaa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	58.5	523	13	BX499268
2	24	58.5	691	10	BE539741
3	24	58.5	713	14	CA788473
4	23.6	57.6	909	29	CNS0521Y

5	23.4	57.1	855	28	BZ141244
6	23.2	56.6	991	29	CG156013
7	23.2	56.6	1018	29	CG431901
8	23.2	56.6	1054	29	CG431789
9	23	56.1	620	12	BJ624928
10	23	56.1	631	12	B096452
11	23	56.1	832	14	CA791463
12	22.8	55.6	517	28	AQ302725
13	22.8	55.6	542	13	BU722189
14	22.8	55.6	725	28	A2896574
15	22.6	55.1	328	10	B224503
16	22.6	55.1	400	13	BY522415
17	22.6	55.1	485	28	B43917
18	22.6	55.1	502	13	BQ104820
19	22.6	55.1	571	13	BX717912
20	22.6	55.1	584	9	AL656843
21	22.6	55.1	597	13	BQ395866
22	22.6	55.1	605	9	AL678053
23	22.6	55.1	623	9	AL677532
24	22.6	55.1	641	9	AL656534
25	22.6	55.1	643	13	BQ390368
26	22.6	55.1	651	9	AL634380
27	22.6	55.1	658	9	AL641573
28	22.6	55.1	660	9	AL662566
29	22.6	55.1	664	9	AL639562
30	22.6	55.1	676	10	BB087044
31	22.6	55.1	681	9	AL658689
32	22.6	55.1	776	14	CF148601
33	22.6	55.1	817	14	CF221876
34	22.6	55.1	861	13	BX715774
35	22.6	55.1	863	13	BX742495
36	22.6	55.1	891	14	CF240425
37	22.6	55.1	904	14	CF241852
38	22.6	55.1	935	14	CF220193
39	22.6	55.1	957	14	CF241149
40	22.6	55.1	1330	28	CC252939
41	22.6	55.1	1408	11	AK034567
42	22.4	54.6	260	12	BP020321
43	22.4	54.6	293	12	BP024827
44	22.4	54.6	304	10	BB269027
45	22.4	54.6	313	13	BW116259

ALIGNMENTS

RESULT 1
BX499268
LOCUS
DEFINITION
523 bp mRNA linear EST 04-SEP-2003
DKFZp779I0544_r1 779 (synonym: hmccl) Homo sapiens cDNA clone
DKFZp779I0544 5', mRNA sequence.
ACCESSION
BX499268
VERSION
EST.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 523)
Bloeker, H., Boecker, M., Mewes, H.W., Weil, B., Amid, C., Oanger, A.,
Fob, G., Han, M. and Wiemann, S.
EST (Bloeker, H., Boecker, M., Mewes, H.W., Weil, B., Amid, C., et al.)
Unpublished (2003)
Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No si sequence available.

BZ141244 CH230-257
CG156013 FUIIP707D
CG431901 Bg49-T7 G
CG431789 Bg110-T3
BJ624928 Bg624928
B096452 B096452
CA791463 AGENCOURT
AQ302725 HS 3217_B
BU722189 SJMAEF03
A2896574 PCJ-24-2
B224503 B224503
BY522415 BY522415
B43917 HS-1058-B1-
BQ104820 fc0556.e
BX717912 BX717912
AL656843 AL656843
BQ395866 NISC dg17
AL678053 AL678053
AL677532 AL677532
AL656534 AL656534
BQ390368 NISC mq12
AL634380 AL634380
AL641573 AL641573
AL662566 AL662566
AL639562 AL639562
BB087044 BB087044
AL658689 AL658689
CF148601 AGENCOURT
CF221876 AGENCOURT
BX715774 BX715774
BX742495 BX742495
CF240425 AGENCOURT
CF241852 AGENCOURT
CF220193 AGENCOURT
CF241149 AGENCOURT
CC252939 CH261-190
AK034567 Mus muscu
BP020321 BP020321
BP024827 BP024827
BB269027 BB269027
BW116259 BW116259

AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Mincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.

TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL Nat. Genet. 25 (2), 235-238 (2000)

MEDLINE 20296633

PUBMED 10835645

REFERENCE 2

AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL Genome Res. 10 (7), 939-949 (2000)

MEDLINE 20359837

PUBMED 10899143

REFERENCE 3 (bases 1 to 909)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES Location/Qualifiers

source 1..909

ORGANISM /organism="Tetraodon nigroviridis"

mol_type /mol_type="genomic DNA"

db_xref /db_xref="taxon:99883"

clone /clone="003H09"

clone lib /clone lib="B"

note /note="Genoscope sequence ID : COAB003CD05C1-end : T7"

ORIGIN

Query Match 57.6%; Score 23.6; DB 29; Length 909;
Best Local Similarity 76.3%; Pred. No. 4.7e+02;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 ATGAACCTACATACACACCTAGTCAGATACACTT 40
|||||

Db 62 ATGAATCATACACACACTAGTAACTAGTCAGATCTCTT 99
|||||

RESULT 5

BZ141244

LOCUS BZ141244 855 bp DNA linear GSS 11-OCT-2002

DEFINITION CH230-257B15.TU CHORI-230 Segment 2 Rattus norvegicus genomic clone

ACCESSION CH230-257B15, genomic survey sequence.

VERSION BZ141244.1 GI:23782191

KEYWORDS GSS.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 855)

AUTHORS Zhao,S., Shetty,J., Shatsman,S., Teegaye,G., Geer,K., Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.

TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment

JOURNAL Unpublished (1999)

COMMENT Other GSSs: CH230-257B15.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Mincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.

TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL Nat. Genet. 25 (2), 235-238 (2000)

MEDLINE 20296633

PUBMED 10835645

REFERENCE 2

AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

JOURNAL Genome Res. 10 (7), 939-949 (2000)

MEDLINE 20359837

PUBMED 10899143

REFERENCE 3 (bases 1 to 909)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES Location/Qualifiers

source 1..909

ORGANISM /organism="Tetraodon nigroviridis"

mol_type /mol_type="genomic DNA"

db_xref /db_xref="taxon:99883"

clone /clone="003H09"

clone lib /clone lib="B"

note /note="Genoscope sequence ID : COAB003CD05C1-end : T7"

ORIGIN

Query Match 57.6%; Score 23.6; DB 29; Length 909;
Best Local Similarity 76.3%; Pred. No. 4.7e+02;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 ATGAACCTACATACACACCTAGTCAGATACACTT 40
|||||

Db 62 ATGAATCATACACACACTAGTAACTAGTCAGATCTCTT 99
|||||

RESULT 6

CG156013

LOCUS CG156013 991 bp DNA linear GSS 21-AUG-2003

DEFINITION PUIIP70TD ZM 0.6-1.0 KB Zea mays genomic clone ZMBTa0591L20, genomic survey sequence.

ACCESSION CG156013

VERSION CG156013.1 GI:34046814

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

REFERENCE 1 (bases 1 to 991)

AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.

TITLE Maize Genomics Consortium

JOURNAL Unpublished (2003)

COMMENT Other GSSs: PUIIP70TB
Contact: Cathy Whitelaw
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.

FEATURES Location/Qualifiers

source 1..991

ORGANISM /organism="Zea mays"

mol_type /mol_type="genomic DNA"

strain /strain="B73"

db_xref /db_xref="taxon:4577"

clone lib /clone lib="ZM 0.6-1.0 KB"

note /note="Vector: PCR4-TOFO; Site 1: EcoRI; 0.6-1.0 kb high Cot selected genomic DNA library"

ORIGIN

Clones are derived from the rat BAC library CHORI-230 (<http://www.chori.org/bacpac/rat230.htm>). For BAC library availability, please contact Pieter de Jong (pdjong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/oreringinformation.htm>). BAC end page: http://www.tigr.org/tdb/bac_end/rat/bac_end_intro.html
Plate: 257 row: B column: 15
Seq primer: SP6
Class: BAC ends.

FEATURES Location/Qualifiers

source 1..855

ORGANISM /organism="Rattus norvegicus"

mol_type /mol_type="genomic DNA"

strain /strain="BN/SENHsd/MCW"

db_xref /db_xref="taxon:10116"

clone /clone="CH230-257B15"

sex /sex="Female"

cell_type /cell_type="Brain"

clone lib /clone lib="CHORI-230 Segment 2"

note /note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI; CHORI-230 Rat (BN/SENHsd/MCW) BAC library produced by Pieter de Jong"

ORIGIN

Query Match 57.1%; Score 23.4; DB 28; Length 855;
Best Local Similarity 73.2%; Pred. No. 5.5e+02;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 GAATGAACCTACATACACACCTAGTCAGATACACTT 41
|||||

Db 636 GAATGAACCTAGAAAACATCATCTCTGAGTGAGTACACATT 676
|||||

RESULT 6

CG156013

LOCUS CG156013 991 bp DNA linear GSS 21-AUG-2003

DEFINITION PUIIP70TD ZM 0.6-1.0 KB Zea mays genomic clone ZMBTa0591L20, genomic survey sequence.

ACCESSION CG156013

VERSION CG156013.1 GI:34046814

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

REFERENCE 1 (bases 1 to 991)

AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.

TITLE Maize Genomics Consortium

JOURNAL Unpublished (2003)

COMMENT Other GSSs: PUIIP70TB
Contact: Cathy Whitelaw
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.

FEATURES Location/Qualifiers

source 1..991

ORGANISM /organism="Zea mays"

mol_type /mol_type="genomic DNA"

strain /strain="B73"

db_xref /db_xref="taxon:4577"

clone lib /clone lib="ZM 0.6-1.0 KB"

note /note="Vector: PCR4-TOFO; Site 1: EcoRI; 0.6-1.0 kb high Cot selected genomic DNA library"

ORIGIN

Eukaryota; Fungi; Glomeromycota; Glomeromycetes; Glomerales;
 Glomeraceae; Glomus.
 1 (bases 1 to 1054)
 Lammers,P.J., Ratyanka,S., Rehner,C. and Jun,J.
 Genomic sequence from Glomus intraradices spore tissue
 Unpublished (2003)
 Contact: Peter Lammers
 New Mexico State University
 BOX 30001, 3MUS, Las Cruces, NM 88003, USA
 Tel: 505-646-3918
 Fax: 505-646-6846
 Email: plammers@nmsu.edu
 Result of blastx search of NCBI non-redundant protein database,
 June 2003: No significant matches.
 Seq primer: T3
 Class: plasmid ends
 High quality sequence stop: 1054.
 Location/Qualifiers
 1. 1054
 /organism="Glomus intraradices"
 /mol_type="Genomic DNA"
 /strain="DAOM 197198, Biosystematic Research Center,
 Ottawa"
 /db_xref="taxon:4876"
 /clone="Bg110-73"
 /tissue_type="Spores"
 /lab_host="E. coli XLR Blue"
 /clone_lib="Glomus intraradices Bgl II library (in pBK-CMV-
 Stratagene)"
 /note="Vector: pBK-CMV; genomic DNA isolated from Spores;
 complete Bgl II digest of gDNA; cloned into BamHI site of
 pBK-CMV"

ORIGIN
 Query Match 56.6%; Score 23.2; DB 29; Length 1054;
 Best Local Similarity 77.8%; Pred.No.6.e+02;
 Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 2 AATGAACACTAATCAACCAACCTTACTGACATACTA 37
 |||||
 692 AATGGACACATACATCACTTGTGCATTAATA 657

RESULT 9
 BJ624928
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 620)
 Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
 Kohara,Y.
 Expressed genes in X. laevis embryo
 Unpublished (2001)
 Contact: Tadasu Shin-i
 Center for Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshinigenes.nig.ac.jp
 The information of this clone is available through the following
 URL.
 http://xenopus.nibb.ac.jp.
 Location/Qualifiers
 1. 620

FEATURES
 SOURCE

```

/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL210a01"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula
library"

ORIGIN
Query Match          56.1%; Score 23; DB 12; Length 620;
Best Local Similarity 74.4%; Pred. No. 7.2e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 AATGAACATACATAACCAACCACTTAGTCAGATCACTACTT 40
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 341 AATGAATAGATGACAAAGACCTTATATAGATCATTAATT 379

RESULT 10
BJ096452          631 bp mRNA linear EST 01-OCT-2003
LOCUS
DEFINITION
Xenopus laevis cdna clone XL154a01 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis (African clawed frog)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 631)
AUTHORS
Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
Expressed genes in X. laevis embryo
Contact: Tadasu Shin-i
Unpublished (2001)
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshinigenes.nig.ac.jp
The information of this clone is available through the following
URL.
http://xenopus.nibb.ac.jp.

FEATURES
source
Location/Qualifiers
1..631
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL154a01"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula
library"

ORIGIN
Query Match          56.1%; Score 23; DB 12; Length 631;
Best Local Similarity 74.4%; Pred. No. 7.2e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 AATGAACATACATAACCAACCACTTAGTCAGATCACTACTT 40
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 550 AATGAATAGATGACAAAGACCTTATATAGATCATTAATT 588

RESULT 11
CA791463          832 bp mRNA linear EST 04-DEC-2002
LOCUS
DEFINITION
AGENCOURT 10304580 NICHG_XGC_Emb1 Xenopus laevis cdna clone
IMAGE:5161647 5', mRNA sequence.

```

```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis (African clawed frog)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 832)
REFERENCE
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11401 row: b column: 16
High quality sequence stop: 417.
FEATURES
source
Location/Qualifiers
1..832
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:5161647"
/tissue_type="embryo (stage 10)"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHG XGC Emb1"
/note="Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI;
Cloned unidirectionally. Primer: Oligo dT. Average insert
size 1.55 kb. Constructed by Life Technologies. Note: This
is a Xenopus Gene Collection (XGC) library."

ORIGIN
Query Match          56.1%; Score 23; DB 14; Length 832;
Best Local Similarity 74.4%; Pred. No. 7.5e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 AATGAACATACATAACCAACCACTTAGTCAGATCACTACTT 40
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153 AATGAATAGATGACAAAGACCTTATATAGATCATTAATT 191

RESULT 12
AQ302725          517 bp DNA linear GSS 16-DEC-1998
LOCUS
DEFINITION
HS 3217 B1_C10 MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3217 Col=19 Row=F, genomic survey
sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 517)
REFERENCE
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
10449764
JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington

```

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 3217 row: F column: 19

Class: BAC ends

High quality sequence stop: 517.

Location/Qualifiers

1. 517

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/clones="Plate:3217 Col=19 Row=F"

/sex="male"

/clone_lib="CIT Approved Human Genomic Sperm Library D"

/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in

E-Coli DH10B"

ORIGIN

Query Match 55.6%; Score 22.8; DB 28; Length 517;

Best Local Similarity 79.4%; Pred. No. 8.2e+02;

Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 7 ACTACATAACACCACTTAGTCAGATACTACTT 40

Db 280 ACGACGTGAAGAACCATGTGAGGCAGATACTACTT 313

RESULT 13

BU722189

LOCUS

DEFINITION BU722189 SJM Schistosoma japonicum cDNA, mRNA sequence. EST 23-OCT-2003

ACCESSION BU722189

VERSION BU722189.1 GI:28329558

KEYWORDS EST

SOURCE Schistosoma japonicum

ORGANISM

Schistosoma japonicum

Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigoida; Schistosomatoidea; Schistosomatidae; Schistosoma.

1. (bases 1 to 542)

Hu, W., Yan, Q., Shen, D.K., Liu, F., Zhu, Z.D., Song, H.D., Xu, X.R.,

Wang, Z.J., Rong, Y.P., Zeng, L.C., Wu, J., Zhang, X., Wang, J.J.,

Xu, X.N., Wang, S.Y., Fu, G., Zhang, X.L., Wang, Z.Q., Brindley, P.J.,

McManus, D.P., Xue, C.L., Feng, Z., Chen, Z. and Han, Z.G.

Evolutionary and biomedical implications of a Schistosoma japonicum

complementary DNA resource

Nat. Genet. 35 (2), 139-147 (2003)

22879925

Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex. 45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn.

Location/Qualifiers

1. 542

/organism="Schistosoma japonicum"

/mol_type="mRNA"

/db_xref="taxon:6182"

/sex="male"

/tissue_type="Whole body"

/dev_stage="adult"

/lab_host="rabbits"

/clone_lib="SJM"

ORIGIN

Query Match

Best Local Similarity 55.6%; Score 22.8; DB 13; Length 542;

Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 6 AACTACATAACACCACTTAGTCAGATACTACT 39

Db 119 ATCTACATAAGAACCCCTAAATAAGATAACT 152

RESULT 14

AZ896574/c

LOCUS

DEFINITION

RPCI-24-209C20-TV RPCI-24 Mus musculus genomic clone

725 bp DNA linear GSS 05-MAR-2001

ACCESSION AZ896574

VERSION AZ896574.1 GI:13215519

KEYWORDS Mus musculus (house mouse)

SOURCE Mus musculus

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1. (bases 1 to 725)

Zhao, S., Nierman, M., Malek, J., Shatsman, S., Akinret, B., Levins, M.,

Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,

Russell, D., de Jong, P. and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-24

Unpublished (1999)

Other GSSs: RPCI-24-209C20.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@email.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.choiri.org/bacpac/orderingframe.htm>). BAC end

page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html

Plate: 209 row: C column: 20

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1. 725

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clones="RPCI-24-209C20"

/sex="Male"

/cell_type="Spleen/Brain"

/clone_lib="RPCI-24"

/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;

RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

ORIGIN

Query Match

Best Local Similarity 55.6%; Score 22.8; DB 28; Length 725;

Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 4 TGAACATACATAACCAACCACTTAGTCAGATACTA 37

Db 702 TGACCTAAATCATACCACTTAGTCAGATGCTA 669

RESULT 15

BB224503

LOCUS

DEFINITION

BB224503 RIKEN full-length enriched, adult male aorta and vein Mus

musculus cDNA clone A530086005.3, mRNA sequence.

ACCESSION BB224503

VERSION BB224503.1 GI:8893115

KEYWORDS EST

SOURCE Mus musculus (house mouse)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 00:16:57 ; Search time 1704 Seconds
(without alignments)
1042.878 Million cell updates/sec

Title: US-09-889-611A-60_COPY_2_42

Perfect score: 41

Sequence: 1 gaatgaactacatacaacc.....ccttagtcagatactacttt 41

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sta.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sta.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_man.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			ID	Description	
		Match	Length	DB			
1	37	90.2	4229	9	AF234618	Homo sapi	
C	2	37	90.2	157284	9	AC072051	Homo sapi
	3	37	90.2	189092	2	AC027524	Homo sapi
	4	21	51.2	6182	6	AX346917	Sequence
C	5	19	46.3	37490	9	AC026356	Homo sapi
	6	19	46.3	74350	8	AP002543	Arabidops
	7	19	46.3	156394	2	AC026558	Homo sapi
C	8	19	46.3	220480	2	AC023973	Homo sapi
	9	18	43.9	2000	6	AX655811	Sequence
	10	18	43.9	143961	2	AP005862	Oryza sat
11	18	43.9	159056	8	CNS08CD7	Oryza sat	
12	18	43.9	215960	2	AC125959	Rattus no	
13	18	43.9	282132	2	AC109976	Rattus no	
C	14	17	41.5	1712	8	AK073122	Oryza sat
	15	17	41.5	16918	6	AX346519	Sequence
	16	17	41.5	107365	4	AC087861	Felis cat
C	17	17	41.5	114149	9	AC022083	Homo sapi
	18	17	41.5	117995	10	AL808138	Mouse DNA
	19	17	41.5	135455	9	AL512666	Human DNA
C	20	17	41.5	148510	10	AC113889	Rattus no
	21	17	41.5	152560	9	AC011339	Homo sapi
	22	17	41.5	157585	2	BX537352	Mus muscu
C	23	17	41.5	162215	10	AL731773	Mouse DNA
	24	17	41.5	162621	2	AC145531	Lemur cat
	25	17	41.5	165764	9	AC116917	Homo sapi
C	26	17	41.5	173933	9	AP002022	Homo sapi
	27	17	41.5	180905	2	AC011688	Homo sapi
	28	17	41.5	188818	2	AC098674	Homo sapi
C	29	17	41.5	189715	2	AC025558	Homo sapi
	30	17	41.5	205202	2	AC139347	Mus muscu
	31	17	41.5	222932	2	AC137869	Mus muscu
C	32	17	41.5	231197	2	AC128935	Rattus no
	33	17	41.5	234545	5	BX470214	Zebrafish
	34	17	41.5	259132	2	AC118381	Rattus no
35	17	41.5	293576	3	AC097830	Rattus no	
36	16	39.0	1766	2	AF070917	Drosophil	
C	37	16	39.0	2076	3	AF429950	Plasmodiu
	38	16	39.0	4116	8	SS1460	patatin A g
	39	16	39.0	6160	6	AX344268	Sequence
C	40	16	39.0	6160	6	AX348665	Sequence
	41	16	39.0	17993	6	AX711896	Sequence
	42	16	39.0	19914	2	AC018009	Drosophil
C	43	16	39.0	37650	9	HSU131B10	Human DNA s
	44	16	39.0	51680	9	AP000261	Homo sapi
	45	16	39.0	54488	8	AF359360	Fusarium

ALIGNMENTS

RESULT 1
AF234618
LOCUS AF234618 Homo sapiens megin gene, promoter region and partial sequence.
DEFINITION AF234618 Homo sapiens megin gene, promoter region and partial sequence.
ACCESSION AF234618.2 GI:19808130
VERSION AF234618.2
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 4229)
AUTHORS Miyata, T., Nangaku, M., Inagi, R. and Kurokawa, K.
TITLE Transcriptional regulation of a mesangium-predominant gene, megin
JOURNAL Unpublished


```

repeat_region      /clone_lib="RPC1-11 Human Male BAC"
repeat_region      599..635
repeat_region      /rpt_family=" (T) n"
repeat_region      1040..1077
repeat_region      /rpt_family="AT_rich"
repeat_region      1411..1747
repeat_region      /rpt_family="L1P"
repeat_region      2632..2992
repeat_region      /rpt_family="MLT1A1"
repeat_region      3115..3138
repeat_region      /rpt_family="AT_rich"
repeat_region      complement(3487..3828)
repeat_region      /rpt_family="L1MC4a"
repeat_region      3834..4338
repeat_region      /rpt_family="MLT1C"
repeat_region      complement(4339..5397)
repeat_region      /rpt_family="L1MC4a"
repeat_region      complement(5742..5899)
repeat_region      /rpt_family="MER5B"
repeat_region      6499..6816
repeat_region      /rpt_family="HAL1"
repeat_region      complement(7416..7621)
repeat_region      /rpt_family="MER3"
repeat_region      complement(8163..9090)
repeat_region      /rpt_family="L2"
repeat_region      9276..9419
repeat_region      /rpt_family="MIR"
repeat_region      9598..9658
repeat_region      /rpt_family="MER81"
repeat_region      9659..10033
repeat_region      /rpt_family="MLT1J"
repeat_region      13986..14009
repeat_region      /rpt_family="AT_rich"
repeat_region      14011..14115
repeat_region      /rpt_family=" (TA) n"
repeat_region      complement(14942..15044)
repeat_region      /rpt_family="MIR3"
repeat_region      15156..15298
repeat_region      /rpt_family="L2"
repeat_region      15592..15773
repeat_region      /rpt_family="L1M4c"
repeat_region      15874..18310
repeat_region      /rpt_family="L1PA16"
repeat_region      16373..16374
unsure
repeat_region      /note="1328 bp of bacterial transposon insertion in BAC
repeat_region      excised between these 2 bp"
repeat_region      18311..18331
repeat_region      /rpt_family=" (TAAAA) n"
repeat_region      18332..19566
repeat_region      /rpt_family="L1PA16"
repeat_region      19567..19587
repeat_region      /rpt_family=" (CA) n"
repeat_region      19588..20042
repeat_region      /rpt_family="L1PA16"
repeat_region      20043..20064
repeat_region      /rpt_family="AT_rich"
repeat_region      complement(21259..21386)
repeat_region      /rpt_family="MIR3"
repeat_region      22395..22420
repeat_region      /rpt_family="AT_rich"
repeat_region      complement(22788..22948)
repeat_region      /rpt_family="FRAM"
repeat_region      complement(23843..24253)
repeat_region      /rpt_family="MSTA"
repeat_region      complement(24678..25041)
repeat_region      /rpt_family="MLT1A2"
repeat_region      25051..25094
repeat_region      /rpt_family="HAL1"
repeat_region      25214..25358
repeat_region      /rpt_family="L1M4c"
repeat_region      complement(25602..25854)
repeat_region      /rpt_family="MIR"
repeat_region      26746..26796

```

```

/rpt_family="AT_rich"
27451..27484
/rpt_family=" (CAAAA) n"
complement(28183..28219)
/rpt_family="MIR"
complement(28249..28325)
/rpt_family="MIR"
complement(29372..29459)
/rpt_family="MIR"
29539..30003
/rpt_family="MER66B"
30596..30624
/rpt_family="AT_rich"
31418..31829
/rpt_family="L2"
31846..31915
/rpt_family=" (TGAA) n"
complement(32019..32103)

```

Query Match 90.2%; Score 37; DB 9; Length 157284;
Best Local Similarity 100.0%; Pred. NO. 3e-12;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATGAACATACATAACCAACACCTTAGTCAGATACTA 37
|||||

Db 155067 GAATGAACATACATAACCAACACCTTAGTCAGATACTA 155031

RESULT 3
AC027524

LOCUS AC027524 189092 bp DNA linear HTG 27-MAR-2003

DEFINITION Homo sapiens chromosome 18 clone RP11-317G1 map 18, 3 unordered pieces.

ACCESSION AC027524 GI:21307437

VERSION AC027524.4

KEYWORDS HTG; HTGS PHASE1; HTGS_FULLTOP; HTGS_CANCELLED.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 189092)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone RP11-317G1

TITLE Unpublished

JOURNAL

AUTHORS

REFERENCE 2 (bases 1 to 189092)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N., Pizani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

TITLE Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

JOURNAL

REFERENCE 3 (bases 1 to 189092)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,

Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., FitzGerald, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kellis, C., Labrecque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneses, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phukhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Sever, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, N., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 1, 2002 this sequence version replaced gi:11990731.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8516

Center clone name: 317_5_1

* NOTE: This is a 'working draft' sequence. It currently

* consists of 3 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 85659: contig of 85659 bp in length

* 85660 85759: gap of 100 bp

* 85760 163014: contig of 77255 bp in length

* 163015 163114: gap of 100 bp

* 163115 189092: contig of 25978 bp in length.

Location/Qualifiers

1. 189092

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosomes="18"

/map="18"

/clone="RP11-317G1"

/clone_lib="RP11-11 Human Male BAC"

ORIGIN

Query Match 90.2%; Score 37; DB 2; Length 189092;

Best Local Similarity 100.0%; Pred. No. 2.9e-12;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATGAACATACATAACACCACTTACTGATGATCTA 37

Db 66563 GAATGAACATACATAACACCACTTACTGATGATCTA 66599

RESULT 4

AX346917/c 6182 bp DNA linear PAT 01-FEB-2002

LOCUS

DEFINITION Sequence 1988 from Patent WO0200928.

ACCESSION AX346917

VERSION AX346917.1 GI:18494803

KEYWORDS

SOURCE synthetic construct

ORGANISM artificial sequences.

REFERENCE

1 Olek, A., Pispembrock, C. and Berlin, K.

AUTHORS

TITLE Diagnosis of diseases associated with the immune system

JOURNAL Patent: WO 0200928-A 1988 03-JAN-2002;

Epigenomics AG (DE)

FEATURES

Location/Qualifiers

1. 6182

/organism="synthetic construct"

/mol_type="unassigned DNA"

/db_xref="taxon:32630"

/note="Chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 51.2%; Score 21; DB 6; Length 6182;

Best Local Similarity 100.0%; Pred. No. 0.038; 0; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 0;

QY 6 AACTACATAACACCACTTCTA 26

|||||

Db 1306 AACTACATAACACCACTTCTA 1286

|||||

RESULT 5

AC026356/c

LOCUS

DEFINITION Homo sapiens 12 BAC RP11-81714 (Roswell Park Cancer Institute Human BAC Library) complete sequence.

ACCESSION AC026356

VERSION AC026356.23 GI:15789206

KEYWORDS

HTG.

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 37490)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbaria, J., Benton, J., Bivaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowles, S., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, I., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louissegh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Marandel, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Merscher, S., Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokoko, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojao, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shim, C.,


```

/evidence=not_experimental
/product="ABC transporter-like protein"
/protein_id="BAB11402.1"
/db_xref="GI:10178109"
/translation="MNQCNVTVSABEDIBAGKKKKPQAEPTLPFLKFRDVTYKVV
KLTSSVEKILGSGSNVGVVGLMPSGSGKTTLLSLAGRISQSSGSGSVYN
DKPYSLKSGIKGVFQDDVLPFLHVTYKTLTAARLRLPKTLTRBQKQALDVTQ
LGLERQDTMGAGFVGVSGGRKRVSGINEIINPSLLLDDEPTSGLDSTALATI
LMLDLIAEAGKVTITTHQPSRLFRFDKLLILGRSLLYFKGSEALDYFSSIGCS
PLTAMPAPKLLDANGINDISVSELDNRVQVNSGRTQTKGSPAAVHEYLVEA
YETRAEQRKLLDVPDDEEAKASTKLRQMGTCWEEQYICLFCGLKRRHRYF
SWLRVTQVLSAVILGLMQSDIRTPMGLDQOGLFTIAVFWGFFVFTAFAPPO
ERAMLNKRAADMYRLSAYFLARTSDLPDLPFLPSLFLVIVFMTGLAISVPYFPLS
MLTVFLCIIIAOQGLAIGAILMDLKKATTLASVTVMTEMLAGGFVFKASPLFLDLC
P"
complement(32784..25652)
/notes="gene_id:F15M7.7"
/codon_start=1
/evidence=not_experimental
/product="selenium-binding protein-like"
/protein_id="BAB11403.1"
/db_xref="GI:10178110"
/translation="MENVILNTRFKPKLALLQSCSPSDUKIHHGFLRLTHLSDV
FVSRLLALCVDDSTFNKPTNLGYAGIFSOIQNPFLNLLRCFSTGAPNSAF
GFYTMKSRIVPDNTITPFLIKASMECVLAGEQTHSQIVRFGQNDVYVENSLSVH
MYANGCFIAAGRPDIPOGFRDVSMTVMVAGYCKGVENAREMDFPHRLFTWS
IMINGYAKNCEKADLPEFKRGVNAVETVMUSVISCAHLGALEGERAYEVV
KSHMTYNLILGALVDMFRCQDIEKAIHVFEGLPTDLSLSSWSSIKGLAVGHAHKA
MYHFSQISGLFPDVTVALVSCSHGLVEKGLIYENMKKHGIEPRLEHYGCI
VMDLGRAGKAEENILKMKVKNAPILGALLGACKIYKTEVAERVGNMLIKVQCI
HPEGYLLSNYACQWQDKIESLRDMKEKLVPPGWSLJEIDGINKFTAGDDQK
HGMGRKRWKEILKRLIGYKNGTDGAPFDVDDEEKSSITHMSEKLAITYAGMMK
TKPGTIRIVKLRVCECHTHTVKLISEVVGRELIVDRNRFFHFRNGVSCRDYH"
complement(26394..27671)
/notes="contains similarity to unknown protein
gb|AAFP71807.1
gene_id:F15M7.8"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11404.1"
/db_xref="GI:10178111"
/translation="MPKCNLLTTSKRRKSKRLKHQHEPSLPPEKEVEEDRDE
GGFKLIAAPSGHEGVQPNLYFNPGVAVNRTNGLNQLILSDEILDILGLGANH
LGVLAIVKTSFYIPANHEPLMNLVLEELKGDPLFNGSKRSTVVAHYHKKFKAGDGE
SNLKIIDVSYDLFQSLCANLEMKVPMKLRNDRITVGRGISVEDFTTKEEPKPVLL
ESLDCWPALEKWSRDLTKVVDVFEVAPVEMKLEKPYRSDGARERPYLFDPK
FAEKVPLSEVDVYFREDLFGVGNRKPDRTWIIIGPAGSGSFHIDPPNSTAWN
AVITGSKWLPFPDVPVPGVHPSPDGAECVPSIIEWFNMFYDDTKDWEKKPTBCI
CKAGEVFPVNGWHLVINLEESIAITQYASR"
complement(join(28281..28383,28556..28602,28740..30146))
/notes="gb|AAFP23201.1
gene_id:F15M7.9
strong similarity to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11405.1"
/db_xref="GI:10178112"
/translation="MDLATVSPTRDVRVRCDCGDCSLNGASPGSLRSVKRKYEEF
ENEKLFHILELIDLSNAKVQLENELELLRETSSQQSQIQDLFELDEERNAASTA
ASRAMNLRQRKABQKLEKQKFAFERKHEHQDQELDLLELITYKREOTIQALT
FQAQAYKHMRSFGFTEAVETEKNNLSRNPMSIENDYQDLPSTDPYPIKCNVNEP
GLEADIDVDVVEKYLADSPHLKTLERRISQMERNPSPFTQPTGDSGGRHYTKNV
VGQSPRQRHFRVRSYGSASLLGTTRKRLDPSNPSRNSGSPRFGKMDPPVAAQNS
FARDKGSSEIGMNDRVYITDSVHSHVSHSGTAEQKPKNDADGYAMSPRELSNQ
PDLGDEISKLYNRLQALDEADRESMQAITSRRTEKXQMKVILKEIAQLHSLKDVVPRR
LPURKTSIIGAFNFSVFKWISFVFWRRKRRSKYMGVQGNMGLQMLLEKTPRIR
QWRCSSTQV"
complement(join(32678..32984,34934..35616))
/notes="gb|AAD04946.2
gene_id:F15M7.10
similar to unknown protein"
/codon_start=1
/evidence=not_experimental

```

```

/protein_id="BAB11406.1"
/db_xref="GI:10178113"
/translation="MGSLEGEFOVADCMGLIQLLNGTVLRSESDILITQQIPFRKN
Query Match 46.3%; Score 19; DB 8; Length 74350;
Best Local Similarity 100.0%; Pred. No. 0.49;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 CACACACCTTAGTCAGATA 34
|||||
Db 26951 CACACACCTTAGTCAGATA 26969
|||||

RESULT 7
AC026558 156394 bp DNA linear HTG 26-MAY-2000
LOCUS Homo sapiens clone RP11-637N6, WORKING DRAFT SEQUENCE, 23 unordered
pieces.
AC026558 3 GI:8077003
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 156394)
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Birken, B., Linton, J., Nusbaum, C. and Lander, E.
Unpublished
REFERENCE 2 (bases 1 to 156394)
AUTHORS Birken, B., Linton, J., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavskiy, L., Boukhgaltier, B., Brown, A., Burkett, G.,
Campoliano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olivarez, T., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7523786.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7470
Center clone name: 637 N 6
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 144229 bases at least Q40
Consensus quality: 150107 bases at least Q30
Consensus quality: 152252 bases at least Q20

```

CDS

CDS

CDS

CDS

Insert size: 163000; agarose-fp
 Insert size: 154194; sum-of-contrigs
 Quality coverage: 4.2 in Q20 bases; agarose-fp
 Quality coverage: 4.4 in Q20 bases; sum-of-contrigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 23 contrigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contrigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1014: contig of 1014 bp in length
* 1015
* 1114: gap of 100 bp
* 1115
2440: contig of 1326 bp in length
* 2441
2540: gap of 100 bp
* 2541
4503: contig of 1962 bp in length
* 4503
4603: gap of 100 bp
* 4603
5920: contig of 1318 bp in length
* 5921
6020: gap of 100 bp
* 6021
8010: contig of 1990 bp in length
* 8011
8110: gap of 100 bp
* 8111
9645: contig of 1535 bp in length
* 9646
9745: gap of 100 bp
* 9746
11265: contig of 1520 bp in length
* 11266
11366: gap of 100 bp
* 11366
12378: contig of 1013 bp in length
* 12379
12478: gap of 100 bp
* 12479
14550: contig of 2072 bp in length
* 14551
14650: gap of 100 bp
* 14651
17260: contig of 2610 bp in length
* 17261
17360: gap of 100 bp
* 17361
19770: contig of 2410 bp in length
* 19771
19870: gap of 100 bp
* 19871
24405: contig of 4535 bp in length
* 24406
24505: gap of 100 bp
* 24506
28007: contig of 3502 bp in length
* 28008
28107: gap of 100 bp
* 28108
32250: contig of 4143 bp in length
* 32251
32350: gap of 100 bp
* 32351
38617: contig of 6267 bp in length
* 38618
38717: gap of 100 bp
* 38718
46909: contig of 8192 bp in length
* 46910
47010: gap of 100 bp
* 47010
56251: contig of 9242 bp in length
* 56252
56351: gap of 100 bp
* 56352
64665: contig of 8314 bp in length
* 64666
64765: gap of 100 bp
* 64766
77790: contig of 13025 bp in length
* 77791
77890: gap of 100 bp
* 77891
88643: contig of 10753 bp in length
* 88644
88743: gap of 100 bp
* 88744
102574: contig of 13831 bp in length
* 102575
102674: gap of 100 bp
* 102675
122629: contig of 19955 bp in length
* 122630
122729: gap of 100 bp
* 122730
156394: contig of 33665 bp in length.

```

FEATURES

```

source
1. .156394
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /clones="RP11-637M6"
  /clone_lib="RP11-11 Human Male BAC"
1. .1014
  /note="assembly_fragment"
1115. 2440
  /note="assembly_fragment"
2541. 4502
  /note="assembly_fragment"
4603. 5920
  /note="assembly_fragment"

```

```

misc_feature 6021. .8010
  /note="assembly_fragment"
misc_feature 8111. .9645
  /note="assembly_fragment"
misc_feature 9746. .11265
  /note="assembly_fragment"
misc_feature 11366. .12378
  /note="assembly_fragment"
  clone_end:SP6
  vector_side:right"
misc_feature 12479. .14550
  /note="assembly_fragment"
misc_feature 14651. .17260
  /note="assembly_fragment"
misc_feature 17361. .19770
  /note="assembly_fragment"
misc_feature 19871. .24405
  /note="assembly_fragment"
misc_feature 24506. .28007
  /note="assembly_fragment"
misc_feature 28108. .32250
  /note="assembly_fragment"
misc_feature 32351. .38617
  /note="assembly_fragment"
misc_feature 38718. .46909
  /note="assembly_fragment"
misc_feature 47010. .56251
  /note="assembly_fragment"
misc_feature 56352. .64665
  /note="assembly_fragment"
misc_feature 64766. .77790
  /note="assembly_fragment"
misc_feature 77891. .88643
  /note="assembly_fragment"
misc_feature 88744. .102574
  /note="assembly_fragment"
misc_feature 102675. .122629
  /note="assembly_fragment"
misc_feature 122730. .156394
  /note="assembly_fragment"
  clone_end:T7
  vector_side:left"

```

ORIGIN

```

Query Match 46.3%; Score 19; DB 2; Length 156394;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 6 AACTACATACACACCT 24

DB 48582 AACTACATACACACCT 48600

RESULT 8

```

AC023973 220480 bp DNA linear HTG 23-SEP-2000
LOCUS Homo sapiens chromosome 12 clone Rp11-793A9 map 12, WORKING DRAFT
DEFINITION SEQUENCE, 28 unordered pieces.
ACCESSION AC023973
VERSION AC023973.4 GI:10280873
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 220480)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,B.
TITLE Homo sapiens chromosome 12, clone Rp11-793A9
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 220480)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,B., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,

```


Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Gande, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Larcocque, K., Lehoczy, J., Levine, R., Liew, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivier, T. M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 23, 2000 this sequence version replaced gi:8076863.
 All repeats were identified using RepeatMasker:
 Smit, A. F. A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L6580
 Center clone name: 793 A_9
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 200376 bases at least Q40
 Consensus quality: 210297 bases at least Q30
 Consensus quality: 214722 bases at least Q20
 Insert size: 234000; agarose-fp
 Insert size: 217780; sum-of-contigs
 Quality coverage: 3.9 in Q20 bases; agarose-fp
 Quality coverage: 4.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 28 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

1
 2203: contig of 2203 bp in length
 2204 2303: gap of 100 bp
 2304 3531: contig of 1228 bp in length
 3532 3632: gap of 100 bp
 3632 4977: contig of 1346 bp in length
 4978 5077: gap of 100 bp
 5078 6822: contig of 1745 bp in length
 6823 9251: contig of 2329 bp in length
 9252 9352: gap of 100 bp
 9352 10776: contig of 1425 bp in length
 10777 14359: contig of 3483 bp in length
 14360 18323: contig of 3864 bp in length
 18324 22156: gap of 100 bp
 22157 22257: contig of 3733 bp in length
 22258 26677: contig of 100 bp
 26678 26777: gap of 100 bp

26778 32091: contig of 5314 bp in length
 32092 32191: gap of 100 bp
 32192 38216: contig of 6025 bp in length
 38217 44697: contig of 6381 bp in length
 44698 50192: contig of 5395 bp in length
 50193 50292: gap of 100 bp
 50293 56609: contig of 6317 bp in length
 56610 65009: contig of 8300 bp in length
 65010 74207: contig of 9098 bp in length
 74208 74307: gap of 100 bp
 74308 81141: contig of 6834 bp in length
 81142 99006: contig of 17765 bp in length
 99007 112940: contig of 13834 bp in length
 112941 125171: contig of 12131 bp in length
 125172 136515: contig of 11244 bp in length
 136516 148849: contig of 12234 bp in length
 148850 161418: contig of 12468 bp in length
 161419 176533: contig of 14916 bp in length
 176534 194705: contig of 18171 bp in length
 194706 210059: contig of 15254 bp in length
 210060 220480: contig of 10322 bp in length
 220481 220480: contig of 10322 bp in length.

FEATURES

source

1. 220480
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="12"
 /map="12"
 /clone="RP11-793A9"
 /clone_lib="RPC1-11 Human Male BAC"

misc_feature

1. 2203
 /note="assembly_fragment"
 clone end:Sp6
 vector side:left
 2304..3531
 /note="assembly_fragment"
 3632..4977
 /note="assembly_fragment"
 5078..6822
 /note="assembly_fragment"
 6923..9251
 /note="assembly_fragment"
 9352..10776
 /note="assembly_fragment"
 10877..14359
 /note="assembly_fragment"
 14460..18323
 /note="assembly_fragment"
 18424..22156
 /note="assembly_fragment"
 22257..26677
 /note="assembly_fragment"
 26778..32091
 /note="assembly_fragment"
 32192..38216
 /note="assembly_fragment"
 38317..44697
 /note="assembly_fragment"
 44798..50192


```

/misc_feature /note="assembly_fragment"
50293. .56609
/note="assembly_fragment"
56710. .65009
/note="assembly_fragment"
65110. .74207
/note="assembly_fragment"
74308. .81141
/note="assembly_fragment"
81242. .99006
/note="assembly_fragment"
99107. .112940
/note="assembly_fragment"
113041. .125171
/note="assembly_fragment"
125272. .136515
/note="assembly_fragment"
136616. .148849
/note="assembly_fragment"
148950. .161417
/note="assembly_fragment"
161518. .176433
/note="assembly_fragment"
176534. .194704
/note="assembly_fragment"
194805. .210058
/note="assembly_fragment"
210159. .220480
/note="assembly_fragment
clone_end:T7
vector_side:right"

ORIGIN
Query Match 46.3%; Score 19; DB 2; Length 220480;
Best Local Similarity 100.0%; Pred.No. 0.42;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 AACTACATACACACCT 24
|||||
Db 117356 AACTACATACACACCT 117374

RESULT 9
AX655811 2000 bp DNA linear PAT 22-MAR-2003
LOCUS Sequence 5681 from Patent WO03000898.
ACCESSION AX655811
VERSION AX655811.1 GI:29158625
KEYWORDS
SOURCE
ORGANISM
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
1
AUTHORS
Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
TITLE
Plant genes involved in defense against pathogens
JOURNAL
Patent: WO 03000898-A 5681 03-JAN-2003;
Syngenta Participations AG (CH)

FEATURES
source
1. .2000
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 43.9%; Score 18; DB 6; Length 2000;
Best Local Similarity 100.0%; Pred.No. 3.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 TACATACACACCTTA 26
|||||

```

```

Db 758 TACATACACACCTTA 775

RESULT 10
AP005862/c 143961 bp DNA linear HTG 31-OCT-2002
LOCUS Oryza sativa (japonica cultivar-group) chromosome 9 clone
DEFINITION OSJNBa0038K02, *** SEQUENCING IN PROGRESS ***.
ACCESSION AP005862
VERSION AP005862.1 GI:24430392
KEYWORDS HTG; HTGS PHASE2.
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
1
AUTHORS
Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 9, BAC
clone-OSJNBa0038K02
JOURNAL
Published Only in Database (2002)
REFERENCE
2 (bases 1 to 143961)
AUTHORS
Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE
Direct Submission
JOURNAL
Submitted (30-OCT-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kamondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(B-mail:tsasaki@ias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1. .143961
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="9"
/clone="OSJNBa0038K02"

ORIGIN
Query Match 43.9%; Score 18; DB 2; Length 143961;
Best Local Similarity 100.0%; Pred.No. 1.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 ACCACCTTAGTCAGATAC 35
|||||
Db 112225 ACCACCTTAGTCAGATAC 112208

RESULT 11
CNS08CD7 159056 bp DNA linear PLN 21-NOV-2003
LOCUS Oryza sativa chromosome 12, . BAC OSJNBb0071117 of library OSJNBb
DEFINITION from chromosome 12 of cultivar Nipponbare of esp. japonica of Oryza
sativa (rice), complete sequence.
ACCESSION AL954853
VERSION AL954853.5 GI:31414507
KEYWORDS HTG.
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
1 (bases 1 to 159056)

```

AUTHORS
 Choiane N., Orjeda G., Cartolico L., Demange N., Wincker P., Segre B., Palletier E., Scarpetti C., Salanoubat M., Weissenbach J. and Quetier F.

TITLE
 Oryza sativa chromosome 12 sequencing

REFERENCE
 2 (bases 1 to 159056)

TITLE
 Direct Submission

JOURNAL
 Submitted (20-NOV-2003) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr)

COMMENT
 - web : www.genoscope.cns.fr
 On Jun 4, 2003 this sequence version replaced gi:28372080.
 Center: Genoscope / Centre National de Sequencage
 Center code: GS
 Web site: <http://www.genoscope.cns.fr/>
 Contact: seque@genoscope.cns.fr

 The following sequence is oriented from the T7 to the SP6 end.
 Upstream BAC (overlapping the T7 end) : OSJNBa0021D06 (AC-BX000560)
 Downstream BAC (overlapping the SP6 end) : OUL618_C05 (AC-AL713952)

 FINISHED SEGMENT STARTS AT BASE 5389
 FINISHED SEGMENT ENDS AT BASE 159056

FEATURES
 source
 1. 159056
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:39947"
 /chromosome="12"
 /clone="OSJNBb0071117"
 /clone_lib="OSJNBb"

ORIGIN
 Query Match 43.9%; Score 18; DB 8; Length 159056;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
 9 TACATAACACACCTTA 26
 |||||
 Db 108806 TACATAACACACCTTA 108823

RESULT 12
 AC125959
 LOCUS
 DEFINITION
 Rattus norvegicus clone CH230-169C9, *** SEQUENCING IN PROGRESS
 *** 24 unordered pieces.
 AC125959 215960 bp DNA linear HTG 26-SEP-2002
 Rattus norvegicus
 AC125959
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 Rattus norvegicus (Norway draft)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

1 (bases 1 to 215960)
 Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M., Hollins, B., Howell, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, J., Jacob, S., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseg, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Maingardne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoelme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Popper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, M., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajic, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission
 Unpublished
 2 (bases 1 to 215960)
 Worley, K. C.

REFERENCE
 Submitted (02-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

TITLE
 Rat Genome Sequencing Consortium.
 3 (bases 1 to 215960)
 Direct Submission
 Submitted (26-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT
 On Sep 26, 2002 this sequence version replaced gi:21671705.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence only reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

 Project Information
 Center project name: GQYN
 Center clone name: CH230-169C9

 Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 159310 bases at least Q40
 Consensus quality: 166804 bases at least Q30
 Consensus quality: 170961 bases at least Q20
 Estimated insert size: 200252; sum-of-contigs estimation
 Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 24 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 31293: contig of 31293 bp in length
 * 31294 31393: gap of unknown length
 * 31394 42669: contig of 11276 bp in length
 * 42670 42769: gap of unknown length
 * 42770 55798: contig of 13029 bp in length
 * 55799 55898: gap of unknown length
 * 55899 62523: contig of 6625 bp in length
 * 62524 62624: gap of unknown length
 * 62624 80012: contig of 17389 bp in length
 * 80013 80112: gap of unknown length
 * 80113 91361: contig of 11249 bp in length
 * 91362 91461: gap of unknown length
 * 91462 171227: contig of 79765 bp in length
 * 171227 171326: gap of unknown length
 * 171327 183652: contig of 12326 bp in length
 * 183653 183752: gap of unknown length
 * 183753 185305: contig of 1553 bp in length
 * 185306 185408: gap of unknown length
 * 185409 185445: contig of 1140 bp in length
 * 185446 186645: gap of unknown length
 * 186646 187846: contig of 1201 bp in length
 * 187847 187946: gap of unknown length
 * 187947 189065: contig of 1119 bp in length
 * 189066 189166: gap of unknown length
 * 189167 190764: contig of 1598 bp in length
 * 190765 190864: gap of unknown length
 * 190865 192694: contig of 1831 bp in length
 * 192695 192794: gap of unknown length
 * 192795 193943: contig of 1148 bp in length
 * 193944 194042: gap of unknown length
 * 194043 196133: contig of 2091 bp in length
 * 196134 196233: gap of unknown length
 * 196234 198533: contig of 2300 bp in length
 * 198534 198633: gap of unknown length
 * 198634 200628: contig of 1995 bp in length
 * 200629 200728: gap of unknown length
 * 200729 202216: contig of 1488 bp in length
 * 202217 202316: gap of unknown length
 * 202317 205020: contig of 2704 bp in length
 * 205021 205120: gap of unknown length
 * 205121 207746: contig of 2626 bp in length
 * 207747 207846: gap of unknown length
 * 207847 209300: contig of 1454 bp in length
 * 209301 209400: gap of unknown length
 * 209401 211460: contig of 2060 bp in length
 * 211461 211560: gap of unknown length
 * 211561 215960: contig of 4400 bp in length.

FEATURES
 source
 1. 215960
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-169C9"
 misc_feature
 21904..24251
 /note="wgs_contig"
 misc_feature
 31394..33471
 /note="wgs_contig"
 misc_feature
 55899..57107
 /note="wgs_contig"
 misc_feature
 80113..82141
 /note="wgs_contig"

ORIGIN

Query Match 43.9%; Score 18; DB 2; Length 215960;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 AATGAACACTACATACAC 19
 |||||
 DB 91069 AATGAACACTACATACAC 91086
 |||||
 RESULT 13
 AC109976 282132 bp DNA linear HTG 26-SEP-2002
 LOCUS Rattus norvegicus clone CH230-177G10, *** SEQUENCING IN PROGRESS
 DEFINITION ***; 25 unordered pieces.
 AC109976
 AC109976.4 GI:23321469
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 282132)
 Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiano,D.,
 Ayalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Faves,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
 Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
 Kowis,C., Kraft,C.I., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
 Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
 Nwokeneme,O., Okwuonu,G., Olarnupsagoon,A., Pal,S., Parks,K.,
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
 Plopper,P., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
 Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
 Sanders,W., Savery,G., Schefer,S., Scott,G., Shatsman,S., Shen,H.,
 Shetty,J., Shvartabeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
 Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
 Williams,G., Willson,R., Wlaczek,R., Wooden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
 Yu,F., Zhang,J., Zhou,J., Zhou,S., Zhao,S., Dunn,D., von
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinstein,G. and Gibbs,R.A.
 Direct Submission

TITLE

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 282132)
Worley, K.C.

Direct Submission
Submitted (09-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 282132)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (26-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Sep 26, 2002 this sequence version replaced gi:21738290.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GQUN
Center clone name: CH230-177G10

----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 215218 bases at least Q40
Consensus quality: 222117 bases at least Q30
Consensus quality: 226410 bases at least Q20
Estimated insert size: 263059; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 64301: contig of 64301 bp in length
* 64302 64401: gap of unknown length
* 64402 79310: contig of 14909 bp in length
* 79311 79410: gap of unknown length
* 79411 90217: contig of 10807 bp in length
* 90218 90317: gap of unknown length
* 90318 105370: contig of 15053 bp in length
* 105371 105470: gap of unknown length
* 105471 108531: contig of 3061 bp in length
* 108532 108631: gap of unknown length
* 108632 127118: contig of 18487 bp in length
* 127119 127218: gap of unknown length
* 127219 153038: contig of 25820 bp in length
* 153039 153138: gap of unknown length
* 153139 182729: contig of 29591 bp in length
* 182730 182829: gap of unknown length
* 182830 193903: contig of 11074 bp in length
* 193904 194003: gap of unknown length
* 194004 206429: contig of 12426 bp in length
* 206430 206529: gap of unknown length
* 206530 207753: contig of 1224 bp in length
* 207754 207853: gap of unknown length
* 207854 209421: contig of 1568 bp in length
* 209422 209521: gap of unknown length
* 209522 211065: contig of 1544 bp in length

* 211066 211165: gap of unknown length
* 211166 212868: contig of 1703 bp in length
* 212869 212968: gap of unknown length
* 212969 216482: contig of 3514 bp in length
* 216483 216582: gap of unknown length
* 216583 218604: contig of 2022 bp in length
* 218605 218704: gap of unknown length
* 218705 221765: contig of 3061 bp in length
* 221766 221865: gap of unknown length
* 221866 225931: contig of 3966 bp in length
* 225932 233131: gap of unknown length
* 233132 233231: contig of 7200 bp in length
* 233232 244973: contig of 11742 bp in length
* 244974 245073: gap of unknown length
* 245074 246570: contig of 1497 bp in length
* 246571 246670: gap of unknown length
* 246671 252670: contig of 6000 bp in length
* 252671 252770: gap of unknown length
* 252771 269281: contig of 16511 bp in length
* 269282 269381: gap of unknown length
* 269382 275653: contig of 6272 bp in length
* 275654 275754: gap of unknown length
* 275755 282132: contig of 6379 bp in length.

FEATURES

source
1..282132
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-177G10"
misc_feature 43744..45689
/note="wgs contig"
misc_feature 64402..68669
/note="wgs contig"
misc_feature 79411..83544
/note="wgs contig"
misc_feature 105471..106633
/note="wgs contig"
misc_feature 107196..108531
/note="wgs contig"
misc_feature 108632..110536
/note="wgs contig"
misc_feature 113702..115613
/note="wgs contig"
misc_feature 115664..116916
/note="wgs contig"
misc_feature 182830..184137
/note="wgs contig"
misc_feature 191871..193903
/note="wgs contig"

ORIGIN

Query Match 43.9%; Score 18; DB 2; Length 282132;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 AATGAACACTACATACAAAC 19
Db 258584 AATGAACACTACATACAAAC 258601

RESULT 14

AK073122/c
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) cDNA clone:J033003A15, full
insert sequence.
ACCESSION
AK073122
KEYWORDS
FLI_CDNA; CAP trapper.
SOURCE
Oryza sativa (japonica cultivar-group)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzae; Oryza.

1
The Rice Full-length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-length cDNA Project Team,
Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K.,
Kishimoto N., Yazaki J., Ishikawa M., Yamada H., Ooka H., Hotta I.,
Kojima K., Namiki T., Ohneda E., Yahagi W., Suzuki K., Li C.,
Ohtsuki K., Shishiki T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group, Ohtsuki Y., Murakami K.,
Iida Y., Sugano S., Fujimura T., Suzuki Y., Taunoda Y.,
Kuroaki T., Kodama T., Masuda H., Kobayashi M., Xie Q., Lu M.,
Narikawa R., Sugiyama A., Mizuno K., Yokomizo S., Niikura J.,
Ikeda R., Ishibiki J., Kawamata M., Yoshimura A., Miura J.,
Kusumegi T., Oka M., Ryu R., Ueda M., Matsubara K., RIKEN,
Kawai J., Carninci P., Adachi J., Aizawa K., Arakawa T., Fukuda S.,
Hara A., Hashidume W., Hayatsu N., Imotani K., Ishii Y., Itoh M.,
Kagawa I., Kondo S., Konno H., Miyazaki A., Oota N., Oca Y.,
Saito R., Sasaki D., Sato K., Shibata K., Shinagawa A., Shiraki T.,
Yoshino M. and Hayashizaki Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice

Science 301 (5631), 376-379 (2003)

22752273

12869764

2 (bases 1 to 1712)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,
Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Inamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Kuroaki, T., Kusumegi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
Namiki, T., Narikawa, R., Niikura, J., Nishii, K., Nomura, K.,
Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,
Oota, N., Ota, Y., Ohtsuki, Y., Ryu, R., Satoh, H., Sakai, C., Sakai, K.,
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and
Yoshimura, A.

Direct Submission

Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp,
Tel: 81-29-838-7007, Fax: 81-29-838-7007)

This clone is one of the 28K full-length cDNA clones from japonica
rice.

URL : <http://cdna01.dna.affrc.go.jp/cDNA/>
NIAS Rice Full-length cDNA Project Team: Kikuchi S., Satoh K.,
Negata T., Kawagashira N., Doi K., Kishimoto N., Yazaki J.,
Ishikawa M., Yamada H., Ooka H., Hotta I., Kojima K., Namiki T.,
Ohneda E., Yahagi W., Suzuki K., Li C., Ohtsuki K., Shishiki T. and
Yamamoto M.

FAIS Genome Sequencing & Analysis Group: Ohtsuki, Y., Iida Y.,
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
Kodama, T., Kuroaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,
Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S.,
Sugiyama, A., Suzuki, Y., Taunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
Yoshimura, A., Matsubara, K. and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,

Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Oota, N.,
Ota, Y., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
Yasunishi, A. and Hayashizaki, Y.

FEATURES
source
1. .1712
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultiVar="Nipponbare"
/db_xref="taxon:39947"
/clone="J033003A15"

ORIGIN
Query Match 41.5%; Score 17; DB 8; Length 1712;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 18 ACCACCTTAGTCAGATA 34
|||||
Db 646 ACCACCTTAGTCAGATA 630
|||||

RESULT 15
AX346519/c
LOCUS 16918 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 1590 from Patent WO0200928.
ACCESSION AX346519
VERSION AX346519.1 GI:18494405
KEYWORDS
synthetic construct
synthetic construct
artificial sequences.

SOURCE
ORGANISM
REFERENCE 1
AUTHORS
TITLE
JOURNAL
EpiGenomics AG (DE)
Patent: WO 0200928-A 1590 03-JAN-2002;

FEATURES
source
1. .16918
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN
Query Match 41.5%; Score 17; DB 6; Length 16918;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 7 ACTACATAACACCCACC 23
|||||
Db 6093 ACTACATAACACCCACC 6077
|||||

Search completed: March 25, 2004, 02:32:49
Job time : 1708 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 00:14:02 ; Search time 249 Seconds
(without alignments)
699.502 Million cell updates/sec

Title: US-09-889-611A-60_COPY_2_42
Perfect score: 41
Sequence: 1 gaatgaactacatacaacc.....ccttagtcagatactacttt 41

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq_29Jan04.*
1: Geneseqm1980s.*
2: Geneseqm1990s.*
3: Geneseqm2000s.*
4: Geneseqm2001as.*
5: Geneseqm2001bs.*
6: Geneseqm2002s.*
7: Geneseqm2003as.*
8: Geneseqm2003bs.*
9: Geneseqm2003cs.*
10: Geneseqm2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	100.0	1431	3	AAA71435 Human meg
2	40	97.6	128	3	AAA71434 Human meg
3	30	73.2	30	3	AAA71449 Human meg
4	30	73.2	30	3	AAA71448 Human meg
5	25	61.0	25	3	AAA71453 Human meg
6	23	56.1	23	3	AAA71452 Human meg
7	21	51.2	6182	6	ABL34015 Human imm
8	19	46.3	1523	3	ACA39949 Arabidops
9	18	43.9	2000	7	ADA72356 Rice gene
10	17	41.5	16918	8	ABL33617 Human DXX
11	17	41.5	50000	8	ADBI6926 Human DXX
12	16	39.0	1652	4	ABL07443 Drosophil
13	16	39.0	3991	6	ABL07442 Drosophil
14	16	39.0	6160	4	ABK31272 Signal tr
15	16	39.0	6160	6	ABL70233 Chemical
16	16	39.0	17993	6	AAD46721 Human tra
17	15	36.6	241	3	AAC13384 Human sec
18	15	36.6	756	4	AAH71525 Human cer
19	15	36.6	789	7	ACA44540 Prokaryot
20	15	36.6	951	7	ACF72327 Staphyloc
21	15	36.6	1002	2	AAT13163 SCSV segm
22	15	36.6	1188	5	AA89552 DNA encod
23	15	36.6	1189	5	AA87858 DNA encod

24	15	36.6	1254	5	AA89553 DNA encod
25	15	36.6	1617	4	AA55391 Haemophil
26	15	36.6	1617	7	ACA34260 Prokaryot
27	15	36.6	1698	4	AAH73359 Human cer
28	15	36.6	1698	5	ABV25202 Human pro
29	15	36.6	1698	5	ABV24391 Human pro
30	15	36.6	1886	3	AA669532 Human sec
31	15	36.6	1886	7	ADA55988 Gene enco
32	15	36.6	1886	7	ADA39801 Human sec
33	15	36.6	1886	7	ACC50434 Human sec
34	15	36.6	1966	6	ABI99599 Mouse isc
35	15	36.6	1986	2	AAQ97797 Clone prd
36	15	36.6	2084	4	ABL10239 Drosophil
37	15	36.6	2560	9	ADB63696 Human cDN
38	15	36.6	5926	6	ABL33653 Human imm
39	15	36.6	6049	6	ABL32228 Human imm
40	15	36.6	6052	4	AA59391 Mouse cDN
41	15	36.6	6075	4	AA59389 Mouse cDN
42	15	36.6	6120	4	AA59390 Mouse cDN
43	15	36.6	6191	6	ABL33216 Human imm
44	15	36.6	6191	6	ABK31306 Signal tr
45	15	36.6	6191	6	ABL70281 Chemical

ALIGNMENTS

RESULT 1
AAA71435
ID AAA71435 standard; DNA; 1431 BP.
XX
AC AAA71435;
XX
DT 01-DEC-2000 (first entry)
XX
DE Human megsin promoter fragment DNA.
XX
KW Promoter; megsin; human; protein isolation; screening. ss.
XX
OS Homo sapiens.
XX
PN WO200043528-A1.
XX
PD 27-JUL-2000.
XX
PF 25-JAN-2000; 2000WO-JP000350.
XX
PR 25-JAN-1999; 99JP-00015667.
XX
PA (KURO/) KUROKAWA K.
PA (MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 2000-543257/49.
XX
PT DNA for promoter region of megsin useful for screening proteins.
XX
PS Disclosure; Fig 2; 45pp; Japanese.
XX
CC This invention describes a novel DNA sequence (I) representing a promoter region having part or all of a specific base sequence. The invention also describes (1) a vector containing (I); (2) a cell transformed by the above vector; and (3) protein produced using (I). (I) is useful for screening and isolating proteins (especially transcription factors). This sequence represents a fragment of the human megsin promoter which is described in the method of the invention

SQ Sequence 1431 BP; 466 A; 266 C; 303 G; 394 T; 0 U; 2 Other;
Query Match 100.0%; Score 41; DB 3; Length 1431;
Best Local Similarity 100.0%; Pred. No. 1.3e-13;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GAATGAACATACATACCAACCACTTAGTCAGATACTACTTT 41
Db 1303 GAATGAACATACATACCAACCACTTAGTCAGATACTACTTT 1343

RESULT 2
AAA71434
ID AAA71434 standard; DNA; 128 BP.
XX AC AAA71434;
XX 01-DEC-2000 (first entry)
XX Human megsin promoter fragment DNA.
XX Promoter; megsin; human; protein isolation; screening. ss.
XX Homo sapiens.
XX WO2000043528-A1.
XX 27-JUL-2000.
XX 25-JAN-2000; 2000WO-JP000350.
XX 25-JAN-1999; 99JP-00015667.
XX (KURO/) KUROKAWA K.
XX (MIYA/) MIYATA T.
XX Miyata T;
XX WPI; 2000-543257/49.
XX DNA for promoter region of megsin useful for screening proteins.
XX Example 5; Page 40; 45pp; Japanese.
XX This invention describes a novel DNA sequence (I) representing a promoter
XX region having part or all of a specific base sequence. The invention also
XX describes (1) a vector containing (I); (2) a cell transformed by the
XX above vector; and (3) protein produced using (I). (I) is useful for
XX screening and isolating proteins (especially transcription factors).
XX AAA71434-A71469 represent PCR primers used in the method described in the
XX invention
XX Sequence 30 BP; 12 A; 9 C; 2 G; 7 T; 0 U; 0 Other;
XX Query Match 73.2%; Score 30; DB 3; Length 30;
XX Best Local Similarity 100.0%; Pred. No. 2.5e-07;
XX Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TACATAACCAACCACTTAGTCAGATACTACTAC 38
Db 1 TACATAACCAACCACTTAGTCAGATACTACTAC 30

RESULT 4
AAA71448
ID AAA71448 standard; DNA; 30 BP.
XX AC AAA71448;
XX 01-DEC-2000 (first entry)
XX Human megsin promoter PCR primer SEQ ID NO: 15.
XX Promoter; megsin; human; protein isolation; screening. PCR primer; ss.
XX Homo sapiens.
XX WO2000043528-A1.
XX 27-JUL-2000.
XX 25-JAN-2000; 2000WO-JP000350.
XX 25-JAN-1999; 99JP-00015667.
XX (KURO/) KUROKAWA K.
XX (MIYA/) MIYATA T.
XX Miyata T;
XX WPI; 2000-543257/49.
XX DNA for promoter region of megsin useful for screening proteins.
XX Example 5; Page 40; 45pp; Japanese.
XX This invention describes a novel DNA sequence (I) representing a promoter
XX region having part or all of a specific base sequence. The invention also
XX describes (1) a vector containing (I); (2) a cell transformed by the
XX above vector; and (3) protein produced using (I). (I) is useful for
XX screening and isolating proteins (especially transcription factors). This
XX sequence represents the human megsin promoter which is described in the
XX method of the invention
XX Sequence 128 BP; 44 A; 29 C; 24 G; 29 T; 0 U; 2 Other;
XX Query Match 97.6%; Score 40; DB 3; Length 128;
XX Best Local Similarity 100.0%; Pred. No. 5.2e-13;
XX Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AATGAACATACATACCAACCACTTAGTCAGATACTACTTT 41
Db 1 AATGAACATACATACCAACCACTTAGTCAGATACTACTTT 40

RESULT 3
AAA71449
ID AAA71449 standard; DNA; 30 BP.
XX AC AAA71449;
XX 01-DEC-2000 (first entry)
XX Human megsin promoter PCR primer SEQ ID NO: 16.
XX Promoter; megsin; human; protein isolation; screening. PCR primer; ss.
XX Homo sapiens.
XX WO2000043528-A1.
XX PN
```

```
CC above vector; and (3) protein produced using (I). (I) is useful for
CC screening and isolating proteins (especially transcription factors).
CC AAA71434-A71469 represent PCR primers used in the method described in the
CC invention
XX
SQ Sequence 30 BP; 13 A; 8 C; 3 G; 6 T; 0 U; 0 Other;

Query Match      73.2%; Score 30; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.5e-07;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 AATGAACATACATAACCAACCACTTAGTCAG 31
DB      1 AATGAACATACATAACCAACCACTTAGTCAG 30

RESULT 5
AAA71453
ID AAA71453 standard; DNA; 25 BP.
XX
AC AAA71453;
XX
DT 01-DEC-2000 (first entry)
XX
DE Human meggin promoter PCR primer SEQ ID NO: 20.
XX
KW Promoter; meggin; human; protein isolation; screening. PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200043528-A1.
XX
DT 27-JUL-2000.
XX
PF 25-JAN-2000; 2000WO-JP000350.
XX
PR 25-JAN-1999; 99JP-00015667.
XX
PA (KURO/) KUROKAWA K.
PA (MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 2000-543257/49.
XX
PT DNA for promoter region of meggin useful for screening proteins.
XX
PS Example 4; Page 42; 45pp; Japanese.
XX
CC This invention describes a novel DNA sequence (I) representing a promoter
CC region having part or all of a specific base sequence. The invention also
CC describes (1) a vector containing (I); (2) a cell transformed by the
CC above vector; and (3) protein produced using (I). (I) is useful for
CC screening and isolating proteins (especially transcription factors).
CC AAA71434-A71469 represent PCR primers used in the method described in the
CC invention
XX
SQ Sequence 25 BP; 8 A; 7 C; 2 G; 8 T; 0 U; 0 Other;

Query Match      61.0%; Score 25; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      17 AACCACTTAGTCAGATACACTTT 41
DB      1 AACCACTTAGTCAGATACACTTT 25

RESULT 6
AAA71452
ID AAA71452 standard; DNA; 23 BP.
XX
AC AAA71452;
```

```
XX
DT 01-DEC-2000 (first entry)
XX
DE Human meggin promoter PCR primer SEQ ID NO: 19.
XX
KW Promoter; meggin; human; protein isolation; screening. PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200043528-A1.
XX
DT 27-JUL-2000.
XX
PF 25-JAN-2000; 2000WO-JP000350.
XX
PR 25-JAN-1999; 99JP-00015667.
XX
PA (KURO/) KUROKAWA K.
PA (MIYA/) MIYATA T.
XX
PI Miyata T;
XX
DR WPI; 2000-543257/49.
XX
PT DNA for promoter region of meggin useful for screening proteins.
XX
PS Example 4; Page 42; 45pp; Japanese.
XX
CC This invention describes a novel DNA sequence (I) representing a promoter
CC region having part or all of a specific base sequence. The invention also
CC describes (1) a vector containing (I); (2) a cell transformed by the
CC above vector; and (3) protein produced using (I). (I) is useful for
CC screening and isolating proteins (especially transcription factors).
CC AAA71434-A71469 represent PCR primers used in the method described in the
CC invention
XX
SQ Sequence 23 BP; 11 A; 7 C; 2 G; 3 T; 0 U; 0 Other;

Query Match      56.1%; Score 23; DB 3; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAATGAACATACATACACACACC 23
DB      1 GAATGAACATACATACACACACC 23

RESULT 7
ABL34015/c
ID ABL34015 standard; DNA; 6182 BP.
XX
AC ABL34015;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1988.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
da.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
```


XX 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX (EPIC-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPI; 2002-130909/17.
 XX
 XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 PT
 XX Claim 1; SEQ ID NO 1988; 32pp + Sequence Listing; German.
 PS
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 CC
 XX Sequence 6182 BP; 1934 A; 37 C; 1059 G; 3152 T; 0 U; 0 Other;
 SQ
 Query Match 51.2%; Score 21; DB 6; Length 6182;
 Best Local Similarity 100.0%; Pred. No. 0.029;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 AACTACATACACACCTTA 26
 Db 1306 AACTACATACACACCTTA 1286
 RESULT 8
 AAC39949/c
 ID AAC39949 standard; DNA; 1523 BP.
 XX
 AC AAC39949;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 26493.
 XX
 XX Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway; metabolic pathway;
 KW promoter; termination sequence; ss.
 XX
 XX Arabidopsis thaliana.
 OS
 PN EF1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-00301439.
 XX
 XX 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 23-APR-1999; 99US-0130891P.
 PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.
 PR 30-APR-1999; 99US-0132407P.
 PR 04-MAY-1999; 99US-0132484P.
 PR 05-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 06-MAY-1999; 99US-0132487P.
 PR 07-MAY-1999; 99US-0132863P.
 PR 11-MAY-1999; 99US-0134256P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134221P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.
 PR 25-MAY-1999; 99US-0136021P.
 PR 27-MAY-1999; 99US-0136392P.
 PR 28-MAY-1999; 99US-0136782P.
 PR 01-JUN-1999; 99US-0137222P.
 PR 03-JUN-1999; 99US-0137528P.
 PR 04-JUN-1999; 99US-0137502P.
 PR 07-JUN-1999; 99US-0137724P.
 PR 08-JUN-1999; 99US-0138094P.
 PR 10-JUN-1999; 99US-0138540P.
 PR 10-JUN-1999; 99US-0138847P.
 PR 14-JUN-1999; 99US-0139119P.
 PR 16-JUN-1999; 99US-0139452P.
 PR 16-JUN-1999; 99US-0139453P.
 PR 17-JUN-1999; 99US-0139492P.
 PR 18-JUN-1999; 99US-0139454P.
 PR 18-JUN-1999; 99US-0139455P.
 PR 18-JUN-1999; 99US-0139456P.
 PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139763P.
 PR 21-JUN-1999; 99US-0139817P.
 PR 21-JUN-1999; 99US-0139899P.
 PR 23-JUN-1999; 99US-0140353P.
 PR 23-JUN-1999; 99US-0140354P.
 PR 24-JUN-1999; 99US-0140695P.
 PR 28-JUN-1999; 99US-0140823P.
 PR 28-JUN-1999; 99US-0140991P.
 PR 30-JUN-1999; 99US-0141287P.
 PR 01-JUL-1999; 99US-0141842P.
 PR 01-JUL-1999; 99US-0142154P.
 PR 02-JUL-1999; 99US-0142055P.
 PR 06-JUL-1999; 99US-0142390P.
 PR 08-JUL-1999; 99US-0142803P.
 PR 09-JUL-1999; 99US-0142920P.
 PR 12-JUL-1999; 99US-0142977P.
 PR 13-JUL-1999; 99US-0143542P.
 PR 14-JUL-1999; 99US-0143624P.
 PR 15-JUL-1999; 99US-0144005P.
 PR 16-JUL-1999; 99US-0144085P.
 PR 16-JUL-1999; 99US-0144086P.
 PR 19-JUL-1999; 99US-0144325P.
 PR 19-JUL-1999; 99US-0144331P.
 PR 19-JUL-1999; 99US-0144332P.
 PR 19-JUL-1999; 99US-0144333P.
 PR 19-JUL-1999; 99US-0144334P.
 PR 19-JUL-1999; 99US-0144335P.
 PR 20-JUL-1999; 99US-0144352P.
 PR 20-JUL-1999; 99US-0144632P.
 PR 20-JUL-1999; 99US-0144884P.
 PR 21-JUL-1999; 99US-0144814P.

```
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0147260P.
PR 02-AUG-1999; 99US-0147303P.
PR 02-AUG-1999; 99US-0147416P.
PR 03-AUG-1999; 99US-0147493P.
PR 03-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149829P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.

PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 46.3%; Score 19; DB 3; Length 1523;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 CAACCACTTAGTCAGATA 34
Db 646 CAACCACTTAGTCAGATA 628

RESULT 9
ADA72356
ID ADA72356 standard; DNA; 2000 BP.
XX AC ADA72356;
XX DT 20-NOV-2003 (first entry)
XX DE Rice gene, SEQ ID 5681.
XX KW Plant; bacterial infection; fungal infection; viral infection; rice;
XX OS gene; ds.
XX OS Oryza sativa.
XX PN WO2003000898-A1.
XX PD 03-JAN-2003.
XX PF 22-JUN-2001; 2001WO-IB001105.
XX PR 22-JUN-2001; 2001WO-IB001105.
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX PA Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
XX PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
XX PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX PS Claim 27; SEQ ID NO 5681; 899pp; English.
XX CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
```

CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.

XX
SQ Sequence 2000 BP; 576 A; 406 C; 423 G; 593 T; 0 U; 2 Other;

Query Match 43.9%; Score 18; DB 7; Length 2000;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 TACATACACACCCCTTA 26
|||||
Db 758 TACATACACACCCCTTA 775

RESULT 10
ABL33617/C
ID ABL33617 standard; DNA; 16918 BP.
XX
AC ABL33617;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1590.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antithyroid; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.

XX Homo sapiens.
XX
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007537.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A. Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX
XX Claim 1; SEQ ID NO 1590; 32pp + Sequence Listing; German.

XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention

XX
SQ Sequence 16918 BP; 4516 A; 235 C; 3886 G; 8281 T; 0 U; 0 Other;

Query Match 41.5%; Score 17; DB 6; Length 16918;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ACTACATACACACCC 23
|||||

Db 6093 ACTACATACACACCC 6077
|||||

RESULT 11
ADB16926
ID ADB16926 standard; DNA; 50000 BP.
XX
AC ADB16926;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human DYXC1 DNA, chromosomal gene region nucleotides 1-50000.
XX
KW ds; human; DYXC1; dyslexia; neurological disorder; chromosome 15q21;
KW reading disability; phonological processing; rapid naming;
KW verbal short-term memory.
XX
XX Homo sapiens.
XX
XX WO2003068814-A1.
XX
XX 21-AUG-2003.
XX
XX 12-FEB-2003; 2003WO-FI000110.
XX
XX 12-FEB-2002; 2002US-0355782P.
XX
XX (LICN) LICENTIA LTD.
XX
XX Kere J, Taipale M, Nopola-Hemmi J, Kaminen N;
XX
XX WPI; 2003-646482/61.
XX
XX New isolated, purified DYXC1 nucleic acid for studying brain processes,
XX e.g. reading, phonological processing, rapid naming or verbal short-term
XX memory, or for diagnosing dyslexia or assessing the predisposition to
XX dyslexia.

XX Claim 27; Page 55-69; 135pp; English.

XX
XX This invention relates to a novel isolated human gene DYXC1 that is
XX functionally related to dyslexia, more particularly it describes single
XX nucleotide polymorphisms thought to predispose an individual in to
XX developing dyslexia. This is a neurological disorder with a genetic basis
XX (DYXC1 has been isolated to chromosome 15q21), which manifests itself as
XX a specific reading disability. Specifically, DYXC1 is can be useful in
XX study of brain processes such as reading, phonological processing, rapid
XX naming and verbal short-term memory. Accordingly, the present invention
XX describes methods and materials for analysing allelic variations in the
XX DYXC1 gene, and also provides DYXC1 as an antigen for the production of
XX antibodies used in the diagnosis of dyslexia. This polynucleotide is the
XX partial genomic sequence of the human DYXC1 chromosomal region
XX (nucleotides 1-50000) of the invention.

XX
SQ Sequence 50000 BP; 15686 A; 9567 C; 10158 G; 14589 T; 0 U; 0 Other;

Query Match 41.5%; Score 17; DB 8; Length 50000;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 CTACATACACACCCCT 24
|||||

Db 24331 CTACATACACACCCCT 24347

RESULT 12
ABL07443
ID ABL07443 standard; cDNA; 1652 BP.
XX
AC ABL07443;
XX
DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 16811.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US009231.
 XX PR 23-MAR-2001; 2000US-0191637P.
 XX PR 11-JUL-2000; 2000US-00614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WIPI; 2001-656860/75.
 XX DR P-PSDB; AB963340.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX PS Claim 1; SEQ ID NO 16811; 21pp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 1652 BP; 434 A; 391 C; 406 G; 421 T; 0 U; 0 Other;
 XX
 Query Match 39.0%; Score 16; DB 4; Length 1652;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 16 CAACCACTTACTGTCAG 31
 Db |||||
 577 CAACCACTTACTGTCAG 592
 RESULT 13
 ABL07442/C
 ID ABL07442 standard; cDNA; 3991 BP.
 XX AC
 XX AC ABL07442;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 16808.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US009231.
 XX PR

PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WIPI; 2001-656860/75.
 XX DR P-PSDB; ABB63339.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX PS Claim 1; SEQ ID NO 16808; 21pp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 3991 BP; 1159 A; 901 C; 865 G; 1066 T; 0 U; 0 Other;
 XX
 Query Match 39.0%; Score 16; DB 4; Length 3991;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 16 CAACCACTTACTGTCAG 31
 Db |||||
 2143 CAACCACTTACTGTCAG 2128
 RESULT 14
 ABR31272/C
 ID ABR31272 standard; DNA; 6160 BP.
 XX AC ABR31272;
 XX DT 23-APR-2002 (first entry)
 XX DE Signal transduction associated gene modified DNA #59.
 XX KW Human; signal transduction associated gene; cytosine methylation state;
 KW CpG island; signal transduction associated disease; solid tumour; cancer;
 KW antitumour; cytostatic; mutant; ds.
 XX OS Homo sapiens.
 XX OS Synthetic.
 XX PN WO200200926-A2.
 XX PD 03-JAN-2002.
 XX PF 29-JUN-2001; 2001WO-EP007472.
 XX PR 30-JUN-2000; 2000DE-01032529.
 XX PR 01-SEP-2000; 2000DE-01043826.
 XX PA (EPIG-) EPIGENOMICS AG.
 XX PI Olek A, Piepenbrock C, Berlin K;
 XX DR WIPI; 2002-147896/19.
 XX PT Oligonucleotide for diagnosis and therapy of diseases associated with
 PT signal transduction e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with signal transduction.

XX Claim 1; SEQ ID NO 115; 24pp; English.
XX
XX The present invention relates to chemically modified DNA sequences of
XX signal transduction associated genes. The DNA sequences are chemically
XX modified using a solution of bisulphite, hydrogen sulphite or disulphite.
XX Also disclosed are oligonucleotides and/or PNA oligomers for detecting
XX the cytosine methylation state (CpG islands) of these genes, and a method
XX for the diagnosis and/or therapy of genetic and epigenetic parameters of
XX genes associated with signal transduction. The genomic DNA can be
XX obtained from cells or cellular components which contain DNA, e.g. cell
XX lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
XX tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
XX brain, heart, prostate, lung, breast or liver, histologic object slides,
XX and all their possible combinations. The sequences of the invention are
XX useful for the diagnosis and therapy of diseases associated with signal
XX transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
XX chemically pretreated genomic DNA sequences of different genes associated
XX with signal transduction, or their complementary sequences. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office
SQ Sequence 6160 BP; 1580 A; 139 C; 1637 G; 2804 T; 0 U; 0 Other;

Query Match 39.0%; Score 16; DB 6; Length 6160;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ACTACATACACACCAC 22
Db |||||||||||||
68 ACTACATACACACCAC 53

RESULT 15
ABL70233/c
ID ABL70233 standard; DNA; 6160 BP.
XX ABL70233;
XX
XX 01-JUL-2002 (first entry)
XX Chemically treated cell signalling DNA sequence#62.
XX
XX Cell signalling; cytosine methylation; cell signalling disease; cancer;
XX tumour; cytostatic; ds.
XX Unidentified.
XX WO200202807-A2.
XX
XX 10-JAN-2002.
XX
XX 29-JUN-2001; 2001WO-EP007471.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-154758/20.
XX
XX Nucleic acid, useful for diagnosis and therapy of diseases associated
XX with cell signaling e.g. cancer, comprises chemically modified genomic
XX sequences of genes associated with cell signaling.

XX Claim 1; SEQ ID NO 123; 24pp + Sequence Listing; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of at least
XX 18 bases of a segment of chemically pretreated DNA of genes associated
XX with cell signalling. The activity of the modified sequences of the

CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office
XX

SQ Sequence 6160 BP; 1580 A; 139 C; 1637 G; 2804 T; 0 U; 0 Other;
Query Match 39.0%; Score 16; DB 6; Length 6160;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ACTACATACACACCAC 22
Db |||||||||||||
68 ACTACATACACACCAC 53

Search completed: March 25, 2004, 02:04:13
Job time : 252 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 01:51:12 ; Search time 2010 Seconds
(without alignments)
609.129 Million cell updates/sec

Title: US-09-889-611A-60_COPY_2_42

Perfect score: 41

Sequence: 1 gaatgaactacataacaacc.....ccttagtcagatactactttt 41

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	46.3	294	10 BF365378	BF365378 QV4-NT002
C 2	18	43.9	549	9 AI188862	AI188862 qd27a12.x
C 3	18	43.9	931	29 CNS02BQ	AI188827 tetracodon
4	17	41.5	72	14 CF608248	CF608248 GEMMA01_0

C	5	17	41.5	228	9	AV319739
C	6	17	41.5	408	28	AO983222
C	7	17	41.5	484	28	AO471300
C	8	17	41.5	531	28	AO807964
C	9	17	41.5	567	28	AZ369484
C	10	17	41.5	569	28	CC162058
C	11	17	41.5	649	14	CF207448
C	12	17	41.5	667	13	BQ997689
C	13	17	41.5	695	13	BQ874693
C	14	17	41.5	782	29	CG114080
C	15	17	41.5	870	13	BX783787
C	16	17	41.5	927	28	CC134385
C	17	17	41.5	1036	10	BE393345
C	18	17	41.5	3764	11	AK087514
C	19	16	39.0	180	14	CD717714
C	20	16	39.0	180	14	CD720798
C	21	16	39.0	259	10	BF067189
C	22	16	39.0	300	9	AV169199
C	23	16	39.0	381	10	BF373680
C	24	16	39.0	397	14	CF404802
C	25	16	39.0	402	9	AV811058
C	26	16	39.0	416	14	CA386931
C	27	16	39.0	427	13	BQ860230
C	28	16	39.0	436	10	AW598133
C	29	16	39.0	438	28	AQ770136
C	30	16	39.0	460	29	FR0041318
C	31	16	39.0	484	10	BQ041997
C	32	16	39.0	485	10	BF425038
C	33	16	39.0	486	14	CF371981
C	34	16	39.0	486	14	CF373282
C	35	16	39.0	513	14	CF405028
C	36	16	39.0	517	10	BE190822
C	37	16	39.0	528	12	BM269742
C	38	16	39.0	533	14	CA936872
C	39	16	39.0	535	12	BM524907
C	40	16	39.0	544	28	AQ993830
C	41	16	39.0	563	28	AQ774541
C	42	16	39.0	564	10	BE347205
C	43	16	39.0	569	9	AI494915
C	44	16	39.0	574	12	BI701328
C	45	16	39.0	576	14	CD799091

ALIGNMENTS

RESULT 1
BF365378/c
LOCUS QV4-NT0028-080700-286-c08 NT0028 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF365378
ACCESSION BF365378.1 GI:11327403
VERSION BF365378.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 294)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,P.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: aaimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV4&t2=QV4-NT0028-080700-286-C08&t3=2000-07-08&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 16
 High quality sequence stop: 294.

FEATURES
 source
 1. .294
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="NT0028"
 /note="Organ: nervous tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN
 Query Match 46.3%; Score 19; DB 10; Length 294;
 Best Local Similarity 100.0%; Pred. No. 5.6;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 AACTACATACACCACT 24
 |||||
 Db 112 AACTACATACACCACT 94

RESULT 2
 A1188862/c
 LOCUS
 DEFINITION qd27a12.x1 Soares_placenta 8to9weeks 2NbHP8to9w Homo sapiens cDNA
 (HUMAN); mRNA sequence.
 A1188862
 A1188862.1 GI:3740071

ACCESSION
 VERSION A1188862.1
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 549)
 NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: ccgaps-remail.nih.gov
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Trace considered overall poor quality
 Insert Length: 1099 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 1.

FEATURES
 source
 1. .549
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:1724926"
 /dev_stage="two placentae: one from 8 weeks and another from 9 weeks post conception"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares_placenta 8to9weeks 2NbHP8to9w"
 /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a

modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo (dT) primer [5' TGTTCACATCTGAGTCGAGCGCGCATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN
 Query Match 43.9%; Score 18; DB 9; Length 549;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 ACCACCTTAGTCAGATAC 35
 |||||
 Db 461 ACCACCTTAGTCAGATAC 444

RESULT 3
 CNS02BQG
 LOCUS
 DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
 253B17 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 A1189827
 A1189827.1 GI:7827931
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetraodon nigroviridis
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontoidea; Tetraodontidae; Tetraodon.

REFERENCE
 1
 Roest Crolius H., Jaillon O., Dasilva C., Bouneau L., Fisher C., Bernot A., Fizames C., Wincker P., Brottier P., Quetier F., Saurin W. and Weissbach J.
 Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
 Nat. Genet. 25 (2), 235-238 (2000)
 20296633
 PUBMED
 10835645

REFERENCE
 2
 Roest Crolius H., Jaillon O., Dasilva C., Ozouf-Costaz C., Fizames C., Fischer C., Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and Weissbach J.
 Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
 Genome Res. 10 (7), 939-949 (2000)
 20359837
 PUBMED
 10899143

REFERENCE
 3 (bases 1 to 931)
 Genoscope.
 Direct Submission
 Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.

FEATURES
 source
 1. .931
 Location/Qualifiers
 /organism="Tetraodon nigroviridis"
 /mol_type="genomic DNA"
 /db_xref="taxon:99883"
 /clone="253B17"
 /clone_lib="G"
 /notes="Genoscope sequence ID : COAG253CA09SP1-end ; PUC-Ori"

ORIGIN


```

QY      1 GAATGAACACTATAACA 17
Db      182 GAATGAACACTATAACA 166

RESULT 6
AQ983222
LOCUS
DEFINITION
  RPCI-23-307F19-TV RPCI-23 Mus musculus genomic clone
  RPCI-23-307F19, genomic survey sequence.
ACCESSION
  AQ983222
VERSION
  AQ983222.1 GI:6816427
KEYWORDS
  GSS.
ORGANISM
  Mus musculus (house mouse)
REFERENCE
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 408)
AUTHORS
  Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S.,
  Akinret, B., Levins, M., McGann, S., Teegaye, G., Geer, K., Krol, M., de
  Jong, P. and Fraser, C.M.
  Mouse BAC End Sequences from Library RPCI-23
  Unpublished (1999)
  Other GSSs: RPCI-23-307F19.TJ
  Contact: Shaying Zhao
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: szhao@tigr.org
  Clones are derived from the mouse BAC library RPCI-23. For BAC
  library availability, please contact Pieter de Jong
  (pieter@jg.med.buffalo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.htm)
  or from Resea ch Genetics (info@resgen.com). BAC end page:
  http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
  Plate: 307 row: F column: 19
  Seq primer: T7
  Class: BAC ends.

FEATURES
  source
  1..408
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /strain="C57BL/6J"
  /db_xref="taxon:10090"
  /clone="RPCI-23-307F19"
  /sex="Female"
  /lab_host="DH10B"
  /clone_lib="RPCI-23"
  /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
  EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
  brain genomic DNA was isolated and partially digested
  with a combination of EcoRI and EcoRI Methylase. Size
  selected DNA was cloned into the pBACe3.6 vector at the
  EcoRI sites. The ligation products were transformed into
  DH10B electrocompetent cells (BRL Life Technologies)."
```

Query Match 41.5%; Score 17; DB 28; Length 408;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      8 CTACATAACACCACT 24
Db      381 CTACATAACACCACT 397

RESULT 7
AQ471300/c
LOCUS
DEFINITION
  AQ471300
  CITBI-E1-2584D24-TR CITBI-E1 Homo sapiens genomic clone 2584D24,
  genomic survey sequence.
ACCESSION
  AQ471300
VERSION
  AQ471300.1 GI:4654954
KEYWORDS
  GSS.
ORGANISM
  Homo sapiens (human)
REFERENCE
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 484)
AUTHORS
  Zhao, S., Adams, M.D., Nierman, W., Malek, J., Shizuya, H., Simon, M. and
  Venter, J.C.
  Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
  Map Building
  Unpublished (1997)
  Other GSSs: CITBI-E1-2584D24.TF
  Contact: Shaying Zhao, William Nierman, Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: hbe@tigr.org
  Clones are available from Research Genetics (info@resgen.com). BAC
  end search page:
  http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
  Seq primer: M13 Reverse
  Class: BAC ends.

FEATURES
  source
  1..484
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /clone="2584D24"
  /sex="male"
  /cell_type="sperm"
  /clone_lib="CITBI-E1"
  /note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
  Caltech Human BAC Library D"

ORIGIN
  Query Match 41.5%; Score 17; DB 28; Length 484;  

  Best Local Similarity 100.0%; Pred. No. 74;  

  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      8 CTACATAACACCACT 24
Db      332 CTACATAACACCACT 316

RESULT 8
AQ807964
LOCUS
DEFINITION
  HS 3155_A2_F03_T7C CIT Approved Human Genomic Sperm Library D Homo
  sapiens genomic clone Plate=3155 Col=6 Row=K, genomic survey
  sequence.
ACCESSION
  AQ807964
VERSION
  AQ807964.1 GI:5727206
KEYWORDS
  GSS.
ORGANISM
  Homo sapiens (human)
REFERENCE
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 531)
AUTHORS
  Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
  Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
  Hood, L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  99380589
  10449764
  Contact: Mahairas GG, Wallace JC, Hood L
  High Throughput Sequencing Center
  University of Washington
```

401 Queen Anna Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@reagen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 3155 row: K column: 6
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 531.

FEATURES source

1. .531
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="Plate=3155 Col=6 Row=K"
 /sex="male"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

ORIGIN

Query Match 41.5%; Score 17; DB 28; Length 531;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CTACATAAACCAACCT 24
 |||||
 Db 460 CTACATAAACCAACCT 476

RESULT 9

AZ369484/c
 LOCUS
 DEFINITION
 1M0120P05F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0120P05 F, genomic survey sequence.

ACCESSION
 AZ369484
 VERSION
 AZ369484.1 GI:10483184
 KEYWORDS
 GSS.

SOURCE
 Mus musculus (house mouse)

ORGANISM

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 567)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,
 Reilly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von
 Niederhausern, A., and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts

JOURNAL

Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA

Tel: 801 585 5606
 Fax: 801 585 7177

Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0120 row: P column: 05

Seq primer: CGTGTAAACGACGGCCAGT
 Class: plasmid ends

High quality sequence stop: 567.

Location/Qualifiers

FEATURES source

1. .567
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0120P05"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"

/clone_lib="Mouse 10kb plasmid UUGC1M library"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 Kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PWD42 (gi|4732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

ORIGIN

Query Match 41.5%; Score 17; DB 28; Length 567;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 TTAGTCAGATCTACTT 40
 |||||
 Db 384 TTAGTCAGATCTACTT 368

RESULT 10

CC162058/c
 LOCUS
 DEFINITION
 i181804.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
 i181804, genomic survey sequence.

ACCESSION
 CC162058
 VERSION
 CC162058.1 GI:30188242
 KEYWORDS
 GSS.

SOURCE
 Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 569)
 Rabinowicz, P.D., O'Shaughnessy, A.L., Ballia, V., Dedhia, N.,
 Katzenburger, F., King, L., Miller, B., Muller, S., Nascentino, L.,
 Zlatavern, T., McCombie, W.R. and Martienssen, R.A.

Genomic shotgun sequences from Zea mays (methyl-filtered)
 Unpublished (2002)

JOURNAL

Contact: W. Richard McCombie
 Lita Annenberg Hazen Genome Sequencing Center
 Cold Spring Harbor Laboratory
 PO Box 100, Cold Spring Harbor, NY 11724, USA
 Tel: 516 367 8884
 Fax: 516 367 8874

Email: mcombie@cshl.org
 Plate: i181 row: b column: 04
 Seq primer: -21M13UnivFwd

Class: shotgun
 High quality sequence stop: 569.

Location/Qualifiers

FEATURES source

1. .569
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone="i181b04"
 /lab_host="DH5a"
 /clone_lib="WGS-ZmaysF (DH5a methyl filtered)"
 /note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;
 The vector was digested with XbaI and one nucleotide was
 added by fill in in the recessive 3' end. The genomic DNA

was nebulized, end repaired, adaptor ligated and size fractionated using sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (.x/y reads in M13mp19, .b/g reads in pUC19). The same ligation was transformed into DH5a."

ORIGIN

Query Match 41.5%; Score 17; DB 28; Length 569;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ACTACATAACAACACC 23
|||||
DB 56 ACTACATAACAACACC 40

RESULT 11

LOCUS CF207448 649 bp mRNA linear EST 01-AUG-2003
DEFINITION CAB20001 IVa_Ra_A09 Cabernet Sauvignon Flower bloom - CAB2 Vitis vinifera cDNA clone CAB20001 IVa_Ra_A09 3', mRNA sequence.
ACCESSION CF207448
VERSION CF207448.1 GI:33401821
KEYWORDS EST
SOURCE Vitis vinifera
ORGANISM Vitis vinifera

REFERENCE

1 (bases 1 to 649)
Goes da Silva, F., Iandolino, A., Lim, H., Baek, J., Jones, K. and Cook, D.
Expressed sequence tags from Vitis vinifera 'Cabernet Sauvignon' berries at various developmental stages
Unpublished (2003)
Contact: Douglas Cook, PhD
CAES Genome Facility
UC Davis, Plant Pathology
One Shields Ave, Davis, CA 95616, USA
Tel: 530 754 6561
Fax: 530 754 6617
Email: drcook@ucdavis.edu
Seq primer: GCCAACGATGCTCTAG.

FEATURES

source
1..649
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CAB20001 IVa_Ra_A09"
/sex="Hermaphrodite"
/dev_stage="Bloom"
/clone_lib="Cabernet Sauvignon Flower bloom - CAB2"
/note="Organ: Flower - Bloom; Vector: pDNR; Site 1: Sfil; Site 2: Sfil; CAB2 is a cDNA library of Vitis vinifera cv. 'Cabernet Sauvignon' Clone 8 berries. Samples were collected at full bloom (80 to 100% flowers showing dehiscence of calyptas or caps and anthers fully extended). Sampled vines were located at the University of California, Davis, Experimental Vineyard. cDNAs were made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
5'-AGCAGTGATCAACGAGTGGCCATTAGCGCGGG-3' and
5'-ATTCAGAGCGGCGGCACATG-dt(30)NN-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."

ORIGIN

Query Match 41.5%; Score 17; DB 14; Length 649;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

6 AACTACATAACAACACC 22

Db 132 AACTACATAACAACACC 148
|||||

RESULT 12

LOCUS BQ997689 667 bp mRNA linear EST 22-AUG-2002
DEFINITION QGG17C07.YG.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone QGG17C07, mRNA sequence.
ACCESSION BQ997689
VERSION BQ997689.1 GI:22432085
KEYWORDS EST
SOURCE Lactuca sativa
ORGANISM Lactuca sativa

REFERENCE

1 (bases 1 to 667)
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenomics.ucdavis.edu/
Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asumundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
singleton, see http://cgdb.ucdavis.edu/ for details.
Plate: QGG17 row: C column: 07.

JOURNAL

COMMENT

Location/Qualifiers
1..667
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L.serriola"
/db_xref="taxon:4236"
/clone="QGG17C07"
/lab_host="E.coli"
/clone_lib="QG_EFGHJ lettuce serriola"
/note="Vector: pBRCNASTAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgdb.ucdavis.edu/
TAG_LIB=QG_EFGHJ lettuce serriola
TAG_SEQ=CGAATGCGGG"

FEATURES

source

Query Match 41.5%; Score 17; DB 13; Length 667;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

QY 16 CAACCCCTTAGTCAGA 32
|||||
DB 163 CAACCCCTTAGTCAGA 179

RESULT 13

LOCUS BQ874693 695 bp mRNA linear EST 15-AUG-2002
DEFINITION QG16A13.YG.ab1 QG_ABCDI lettuce salinas Lactuca sativa cDNA clone QG16A13, mRNA sequence.
ACCESSION BQ874693

```

VERSION      BQ874693.1  GI:22261253
KEYWORDS     EST.
SOURCE       Lactuca sativa
ORGANISM     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
              asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
              Cichorieae; Lactuca.
REFERENCE    1 (bases 1 to 695)
AUTHORS      Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
              Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,
              Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
              Lai,Z., Church,S., Jackson,L. and Bradford,K.
              Lettuce and Sunflower ESTs from the Compositae Genome Project
              http://compgenomics.ucdavis.edu/
TITLE        Unpublished (2002)
JOURNAL      Department of Vegetable Crops, R.W.Michelmore Lab
COMMENT      University of California at Davis (UCD)
              Asmundson Hall, UCD, Davis, CA 95616, USA
              Tel: 1-(530)-742-1742
              Fax: 1-(530)-752-9659
              Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]
              singleton, see http://cgdb.ucdavis.edu/ for details.
              Plate: QG16 row: A column: 13.
FEATURES     Location/Qualifiers
              source
                1..695
                /organism="Lactuca sativa"
                /mol_type="mRNA"
                /cultivar="Salinas"
                /db_xref="taxon:4236"
                /clone="QG16A13"
                /lab_host="E.coli"
                /clone_lib="QG ABCDI lettuce salinas"
                /note="Vector: pBRCNDSFIAB; The library was constructed
              from 10 different sources of RNA from a single genotype.
              Separate cDNAs were generated using primers that
              incorporated unique 5' and 3' tags to distinguish each
              source of RNA. cDNAs were then pooled, size-fractionated,
              directionally cloned into a custom medium-copy vector and
              transformations made with four size classes to minimize
              size bias. Details of each source of RNA and library
              construction can be obtained at http://cgdb.ucdavis.edu/
              TAG_TISSUE=roots environmental stress
              TAG_LIB=QG ABCDI lettuce salinas
              TAG_SEQ=ATCTCGCGG"

ORIGIN
Query Match      41.5%; Score 17; DB 13; Length 695;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ATGAACCTACATACAAAC 19
    |||||
Db 205 ATGAACCTACATACAAAC 221

RESULT 14
CG114080/0
LOCUS      CG114080
DEFINITION PUJGG81TD ZM.0.6.1.0 KB Zea mays genomic clone ZMMBTa0661N17,
            genomic survey sequence.
ACCESSION  CG114080
VERSION     CG114080.1  GI:33997517
KEYWORDS    GSS.
SOURCE      Zea mays
            Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 782)
            Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and

Bennetzen,J.
Maize Genomics Consortium
Unpublished (2003)
Other GSSs: PUJGG81TB
Contact: Cathy Whitelaw
TIGR Medical Center Drive, Rockville, MD 20850, USA
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
FEATURES     Location/Qualifiers
              source
                1..782
                /organism="Zea mays"
                /mol_type="genomic DNA"
                /strain="B73"
                /db_xref="taxon:4577"
                /clone="ZMMBTa0661N17"
                /clone_lib="ZM 0.6.1.0 KB"
                /note="Vector: PCR4-TOP0; Site 1: EcoRI; 0.6-1.0 kb high
              Cot selected genomic DNA library"

ORIGIN
Query Match      41.5%; Score 17; DB 29; Length 782;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ACTACATAACACCCACC 23
    |||||
Db 497 ACTACATAACACCCACC 481

RESULT 15
BX783787
LOCUS      BX783787
DEFINITION BX783787 XGC-egg Silurana tropicalis cDNA clone Tegg059102 3', mRNA
            sequence.
ACCESSION  BX783787
VERSION     BX783787.1  GI:39690993
KEYWORDS    EST.
SOURCE      Silurana tropicalis (western clawed frog)
            Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
            Xenopodinae; Silurana.
            1 (bases 1 to 870)
            Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
            Sanger Xenopus tropicalis EST project 2001 (11_2003)
            Unpublished (2003)
            Contact: Croning MDR
            Sanger Institute
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: trop@sanger.ac.uk
            TROPICALIS_SEQUENCE_ID: Tegg059102.q1kx77
            Sequencing primer: T7
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Aaron M. Zorn.
            cDNA was oligo dt primed from sug of poly A+ RNA from egg.
            EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
            5' end and NotI at the 3' end.
            Vector: pCS107; Site 1: EcoRI; Site 2: NotI
            Host: Escherichia coli XL1-blue.
            Location/Qualifiers
              source
                1..870
                /organism="Silurana tropicalis"
                /mol_type="mRNA"
                /db_xref="taxon:8364"
                /clone="Tegg059102"
                /dev_stage="egg"
                /lab_host="Escherichia coli XL1-blue"
                /clone_lib="XGC-egg"

```

/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"

ORIGIN

Query Match 41.5%; Score 17; DB 13; Length 870;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 21 ACCTTAGTCAGATACTA 37

Db 287 ACCTTAGTCAGATACTA 303

Search completed: March 25, 2004, 03:06:30
Job time : 2014 secs